UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DEPARTMENT OF INFORMATION ENGINEERING

Master of Science in
Communication Technologies and Multimedia

MASTER THESIS

Adaptive loudness compensation in audio reproduction

Supervisor:
Prof. Pierangelo Migliorati

Co-supervisors:
Prof. Vesa Välimäki
Dr. Jussi Rämö

Candidate:
Leonardo Fierro

Candidate number:
704495

Academic Year 2017/2018

UNIVERSITÀ DEGLI STUDI DI BRESCIA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Corso di Laurea Magistrale in
Communication Technologies and Multimedia

TESI DI LAUREA MAGISTRALE

Compensazione adattiva del volume nella riproduzione audio

Relatore:
Prof. Pierangelo Migliorati

Correlatori:
Prof. Vesa Välimäki
Dr. Jussi Rämö

Candidato:
Leonardo Fierro

Matricola:
704495

Anno Accademico 2017/2018

See you, space cowboy.

Somewhere, somehow.

Acknowledgements

I would like to thank all the people that helped and supported me during this thesis work and, more in general, the entire master period.

To begin with, I am grateful to my thesis supervisor, Prof. Pierangelo Migliorati, for the availability and assistance he gave me during my research and for the opportunity to work abroad on such a particular and interesting topic.

Secondly, I want to sincerely thank Prof. Vesa Välimäki and Dr. Jussi Rämö of the Aalto Acoustics Lab at Aalto University, Espoo, Finland. They welcomed me warmly and supervised my work, transmitting passion and dedication and stimulating my personal growth. I would like to extend my thanks to all the members of the Acoustics Lab staff, who agreed to be involved in my validation test, as well.

Finally, I must express my profound gratitude and love to my family, always supporting me unconditionally and pushing me to give my best. I am so proud of them, I hope that today they feel a little proud of me too.

The same gratitude I must show to all those friends I shared this journey with: first of all, the "Scansadiavoli", who have been by my side for so many years now that I can not help but consider them as family; Manuel, my trusty companion for music, laughter and nerd nonsense; my team-mates from Rugby Brescia; and, last but not least, my CTMLM classmates, with whom I shared moments of joy, passionate discussions, astonishing frags and, probably, way too many cups of coffee.

All the people mentioned on this page have played a fundamental role in the preparation of this thesis; unfortunately, every mistake or inaccuracy is attributable only to me.

- Leonardo

UNIVERSITY OF BRESCIA
Department of Information Engineering

Abstract

Adaptive loudness compensation in audio reproduction

by Leonardo Fierro

This work involves the study of the psychoacoustic phenomenon of nonlinear and frequency dependent loudness perception, its modeling, and the use of digital filters to introduce an adaptive compensation based on the reproduction level.

Music and sound are mixed and mastered at a particular loudness level, which is usually louder than the level they are commonly played at. This implies a change in the perceived spectral balance of the audio source, which is largest in the bass and sub-bass ranges. As the volume setting in music reproduction is decreased, a loudness compensation filter can be designed to introduce an appropriate boost, so that the low frequencies are still heard well and the perceived spectral balance is preserved.

This thesis describes a loudness compensation function derived from the standard *equal-loudness level contours* and its implementation via a digital first-order shelving filter, and it documents a formal listening test, designed and conducted to validate the accuracy of such a method.

The research work was carried out between October 2018 and January 2019, during a visiting period at the Aalto University, Espoo, Finland.

UNIVERSITÀ DEGLI STUDI DI BRESCIA
Dipartimento di Ingegneria dell'Informazione

Compendio

Compensazione adattiva del volume nella riproduzione audio

di Leonardo Fierro

Questo lavoro comprende lo studio del fenomeno psicoacustico della percezione non lineare e dipendente dalla frequenza della rumorosità del suono (meglio definita come *loudness*) e la sua modellizzazione, utilizzando filtri digitali per realizzare una compensazione adattiva basata sul livello di ascolto.

Musica e suoni sono missati e masterizzati ad un certo livello di volume, tipicamente più alto rispetto a quello al quale sono comunemente riprodotti. Questo implica una variazione nel bilanciamento spettrale, molto evidente nell'intervallo di basse frequenze. Mentre il volume della musica in ascolto viene ridotto, è possibile introdurre un'appropriata compensazione dei bassi tramite dei filtri, in modo che le basse frequenze siano rese nuovamente udibili e che il bilanciamento spettrale sia ripristinato.

In questa tesi è descritta una funzione di compensazione della loudness derivata dalle *curve di egual livello* standard e un'implementazione della stessa tramite filtri a scaffale del primo ordine. L'accuratezza del metodo proposto è validata dai risultati ottenuti da un test di ascolto formale.

Il lavoro di ricerca è stato svolto tra Ottobre 2018 e Gennaio 2019, durante un periodo di visita presso l'Università di Aalto, ad Espoo, Finlandia.

Contents

List of Figures

List of Tables

List of Abbreviations

L_M	**M**astering **L**evel
L_L	**L**istening **L**evel
ELLC	**E**qual **L**oudness **L**evel **C**ontours
SPL	**S**ound **P**ressure **L**evel
MAF	**M**inimum **A**udible **F**ield
dBA	**A**-weighted **d**eci**b**el
LKFS	**L**oudness, **K**-weighted, relative to **F**ull **S**cale
IIR	**I**nfinite **I**mpulse **R**esponse
FIR	**F**inite **I**mpulse **R**esponse
SNR	**S**ignal-to-**N**oise **R**atio
FOF	**F**ractional **O**rder **F**ilter
GA	**G**enetic **A**lgorithm
IQR	**I**nter-**Q**uartile **R**ange

Introduction

The need for loudness compensation has been acknowledged since the first documented equal-loudness level contours (ELLC) reported by Fletcher and Munson [5]; different approaches, both in the analog [8] [18] [28] and in the digital universe [7] [25] [30], have been discussed and proposed since then. It is well known how perceived bass and sub-bass ranges are much more affected than high frequencies when the audio reproduction level decreases. As a consequence, it is valuable to be able to adapt the compensation based on the listening level of the playing track.

According to Katz [16], music is nowadays usually mixed and mastered with a set of loudspeakers at the sound pressure level (SPL) of 83 dB, or more generally at levels between 80 to 85 dB. In that range, human loudness perception is the closest to be flat while avoiding painful listening conditions. However, those levels are quite high: a prolonged exposure can tire the listener or even damage the hearing [20] [38]. Safer listening levels for consumers (in particular using headphones) lie in the 60 to 75 dB SPL interval.

Inevitably, when the sound reproduction level is changed, the perceived spectral balance is altered as well and the fidelity to the original master is lost. Therefore, the ultimate goal of a loudness compensation method is not to provide the best subjective bass compensation according to the listener, but to recover that lost fidelity by regaining the spectral balance of the playback sound.

Consumer audio equipment sometimes offer a "loudness compensation" feature, whose action is merely a +10 dB bass boost regardless of the playback level [8].

This thesis work, instead, describes a compensation technique using low-order digital filters to improve the listening experience, based on the equal-loudness contours provided by the ISO 226:2003 standard [12].

This document is organized as follows. Chapter 1 provides some background theory on the human hearing system and briefly illustrates the equal-loudness level contours, with their evolution and refinement through time. Chapter 2 describes the compensation method, starting from the standard data interpolation. Chapter 3 is related to the filter design and the optimization of filter parameters. Description and results of the conducted listening test are shown in Chapter 4. Appendixes describing valuable parts of code implementations are reported at the end of thesis.

Chapter 1

Human ear and loudness perception

In this first chapter, some basic theory regarding psychoacoustics is presented, in order to provide a minimum background for the later-proposed compensation method. In particular, the main anatomical structures of the hearing system and the human perception of loudness are described, together with a small section relative to the design process of a listening test.

1.1 Psychoacoustics

Psychoacoustics is "the study of sound perception and audiology" [3]. In particular, it deals with the psychological and physiological responses associated with sound (whether it is music, speech or noise) and the relationship between psychology (perception) and physics (physical variables) [23].

The origin of psychoacoustic studies is set in the greek era, when Pythagoras first attempted to understand the physical and mathematical bases of musical scales. Deeper knowledge of the relationship between frequency of vibration and pitch came with the works of Leonardo Da Vinci (16th century), Galileo Galilei (17th century) and Felix Savart (19th century). However, the leading scientist and field trailblazer was Hermann von Helmholtz (1821-1894): his book *"On the Sensations of Tone as a Physiological Basis for the Theory of Music"* had been the major reference for hearing and musical perception for many decades [40].

Psychoacoustics has then evolved and deepened through time, and it is currently a very active field of research. Some of the main topics are, for example, musical tuning, auditory scene analysis, speech recognition, sound masking, sound localization and pitch and timbre determination.

1.2 Anatomy of the human ear

The human hearing system can be anatomically divided into three main sections:

- *outer* or *external ear*, comprising the pinna, the concha and the auditory canal;
- *middle ear*, comprising eardrum and ossicles;
- *inner ear*, comprising the cochlea and the vestibular system.

Anatomical structure of the ear and its simplified scheme can be found in Fig. 1.1 and 1.2.

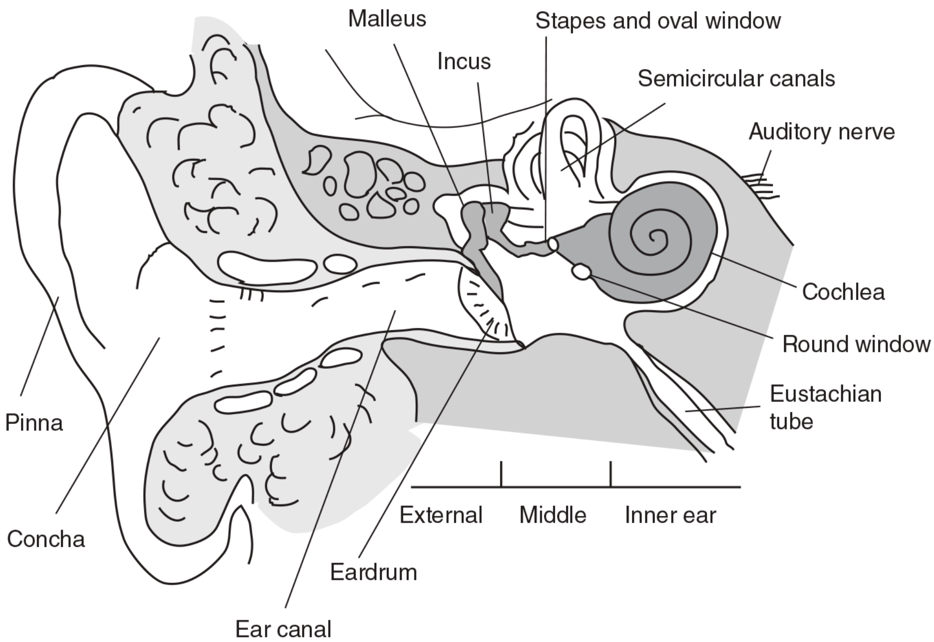


FIGURE 1.1: Anatomical structure of the human ear.

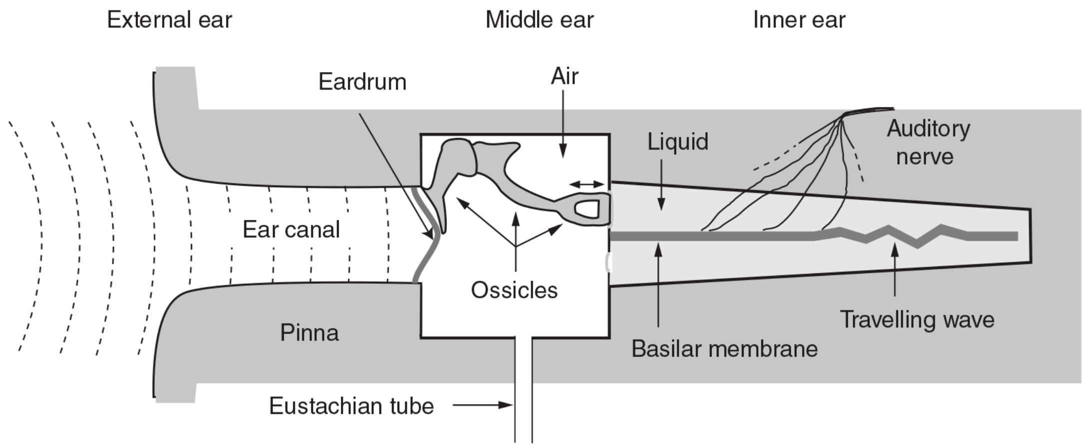


FIGURE 1.2: Simplified structure of the human ear.

1.2.1 Outer ear

The outer ear (*auris externa*) consists mainly of the auricle (commonly known as "pinna"), the concha and the auditory canal up to the tympanic membrane. Its main function is to channel the incoming sound waves inside the auditory canal, gathering sound energy and focusing it on the tympanic membrane.

The pinna is the visible part of the ear. It is composed of a thin plate of elastic cartilage, covered with integument and connected to the surrounding parts by ligaments and muscles and to the ear canal by fibrous tissue. Its shape, with its rigs and its depressions, helps sound localization [26].

The concha is the entrance of the auditory canal and acts as an acoustic resonant cavity. The canal is typically 25-to-35 mm long and conducts sound vibrations to the tympanic membrane.

The combined acoustic effect of the outer ear as a whole (Fig. 1.3) changes the frequency response of incoming sounds waves due to resonance effects. Outer ear configuration selectively boost frequencies around 3 kHz, where human speech sounds are mainly located. This amplification also makes humans most sensitive to frequencies in such an interval and so explains why they are particularly prone to acoustical injury and hearing loss near that frequency [27].

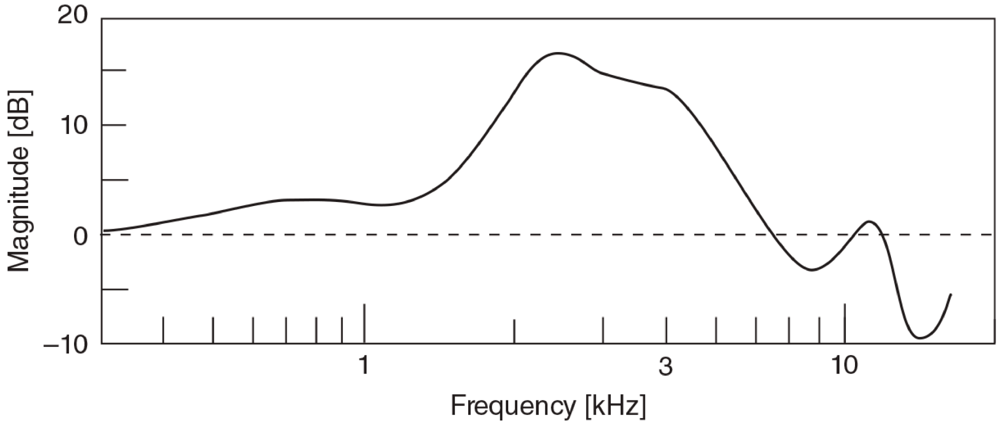


FIGURE 1.3: Approximate magnitude response of the outer ear at the eardrum to a frontal sound source in free-field condition.

The tympanic membrane (also called *eardrum*) is a thin and elastic membrane forming a boundary interface between the outer ear and the middle ear (Fig. 1.4). It converts acoustic pressure variations from the outside world into mechanical vibrations in the middle ear. Breaking or perforating the eardrum can lead to serious hearing loss.

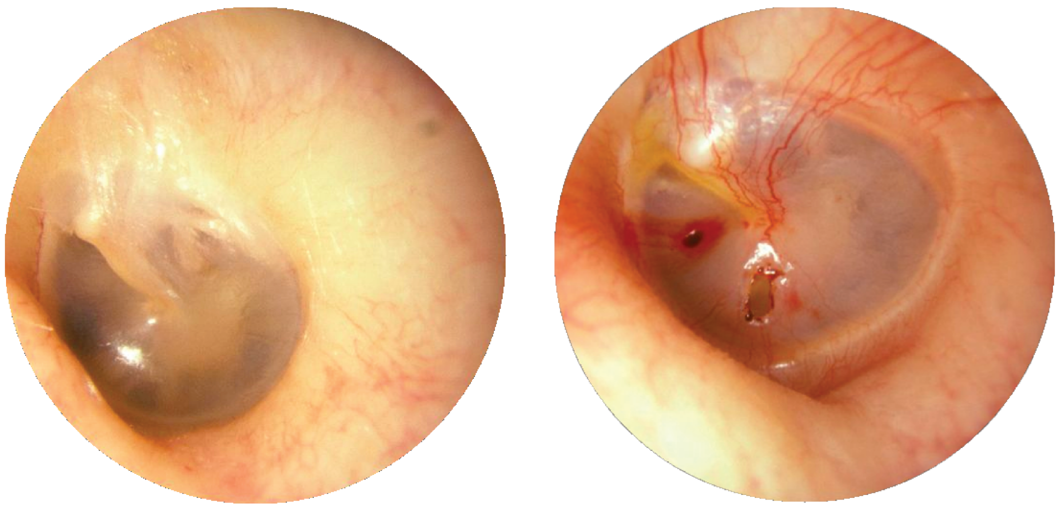


FIGURE 1.4: Examples of a normal tympanic membrane (left) and a ruptured one (right).

1.2.2 Middle ear

Mechanical vibrations produced by the eardrum are passed and transferred into waves in the fluid and membranes of the inner ear through a sequence of three small bones called *ossicles*, comprising the malleus ("hammer"), incus ("anvil") and stapes ("stirrup"). The stapes footplate is attached to the oval window of the cochlea, which acts as the boundary interface between the middle ear and the inner ear [17].

The malleus is fixed to the middle fibrous layer of the tympanic membrane and strongly joined to the incus so that, at normal sound intensity levels, they act as a single unit rotating together as the tympanic membrane vibrates to move the stapes via a ball-and-socket joint, as a piston [9]. The hollow space of the middle ear is known as the tympanic cavity and is surrounded by the tympani bone. The Eustachian tube joins the tympanic cavity with the nasal cavity, allowing pressure equalization between the middle ear and the throat.

The two main functions of the middle ear are:

- efficient transmission of vibrations from the eardrum to the fluid that fills the cochlea;
- protection of the hearing system from the impact of loud sounds.

The ossicles enable a mechanical impedance matching [17], overcoming the higher resistance to movement of the cochlear fluid compared with the one of the air at ear entrance (Fig. 1.6).

1.2.3 Inner ear

The inner ear (*auris interna*) is the region mainly responsible for balance and sound detection [35]; it consists of the bony labyrinth, a hollow cavity in the temporal bone of the skull, comprising two principal functional parts [39]:

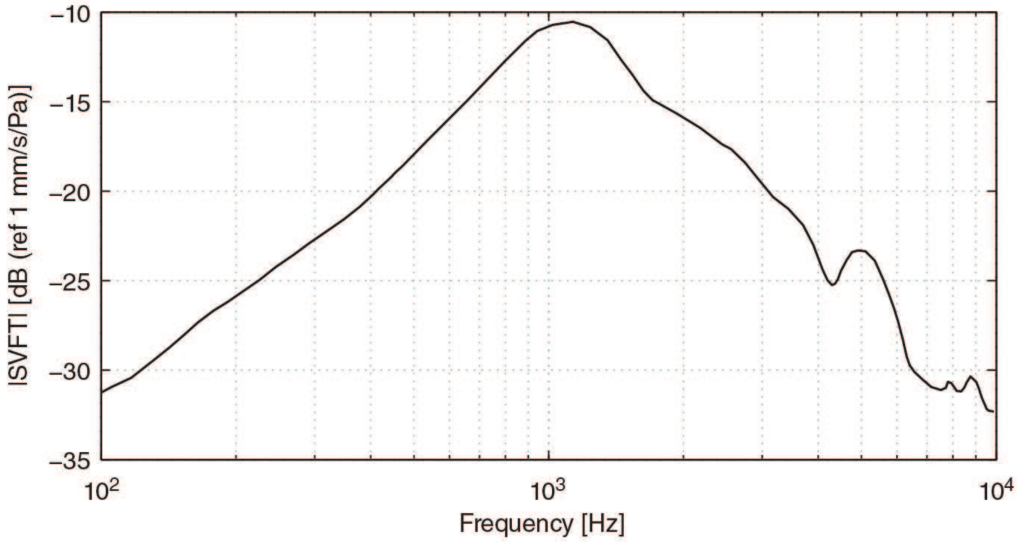


FIGURE 1.5: Middle ear pressure transfer function: temporary superior vena cava filter.

- the cochlea, where mechanical vibration coming from the middle ear are converted into electro-chemical impulses (*firings*), passed onto the brain via the auditory nerve;
- the vestibular system, dedicated to balance.

The cochlea is essentially a spiral tube with two ends, one closed ("apex") and the other ("base") with two openings: the oval window and the round window. The

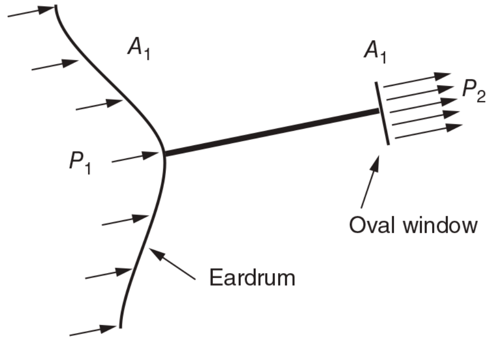


FIGURE 1.6: Impedance matching in the middle ear.

tube is split into three sections by Reissner's membrane and the basilar membrane (Fig. 1.7). The inner channel is called *scala media*; the outer channels, *scala vestibuli* and *scala tympani*, are filled with an uncompressible fluid, called perilymph. The scala vestibuli terminates at the oval window, the scala tympani at the round window. There is a small hole at the apex known as *helicotrema*, through which the perilymph fluid can flow and close the loop.

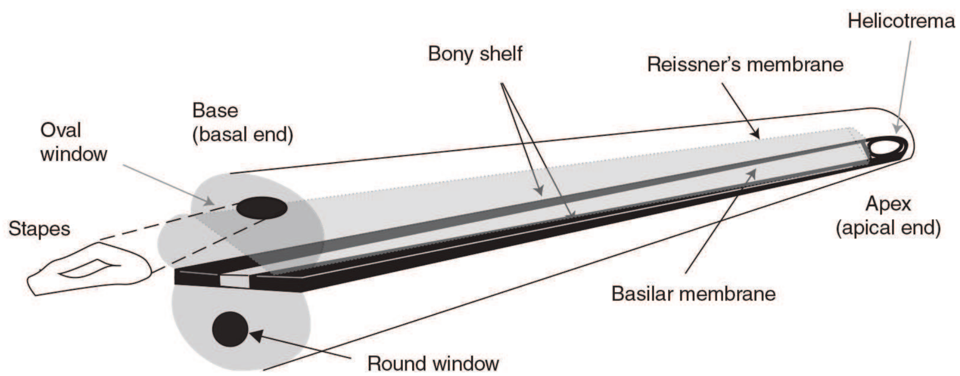


FIGURE 1.7: Schematic of the cochlea, depicted as a linear tube instead of its true, spiral form.

Input acoustic vibrations result in a piston-like movement of the stapes footplate at the oval window, which moves the perilymph fluid within the cochlea and, as a consequence, the round window membrane to compensate. These movements set travelling waves in the scala vestibuli, where the basilar membrane carries out a frequency analysis of input sounds [9].

Basilar membrane is narrow and thin at the base and becomes wider and thicker along its length to the apex. When stimulated by audio signals, different regions vibrate along the membrane (Fig. 1.8): best response to high frequencies happens at the base, to low frequencies at the apex. Since membrane thickness and width change gradually, input pure tones at different frequencies will produce a maximum of vibration at different *places*, i.e. different positions.

The point of maximum displacement along the basilar membrane changes as the frequency of the input is altered (Fig. 1.9): the linear distance from the apex to such

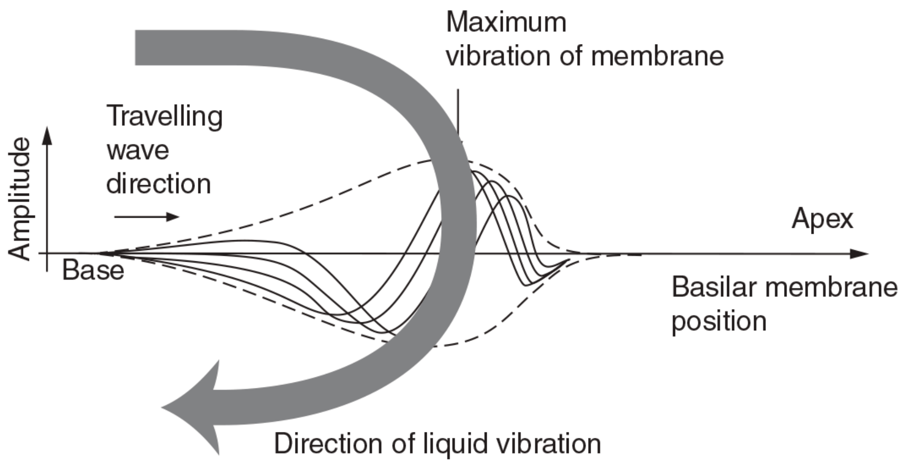


FIGURE 1.8: Displacement of the travelling wave on the basilar membrane.

a point is directly proportional to the logarithm of the frequency.

Basilar membrane movements are then converted into nerve firings by the organ of Corti, in order to be transmitted to the brain [9]. The organ of Corti is located on the basilar membrane (Fig. 1.10) and contains around 15.500 hair cells (around 3.500 inner hair cells and 12.000 outer hair cells), which present *stereocilia*, groups of 10-to-50 μm long hair (or *cilia*), at their end.

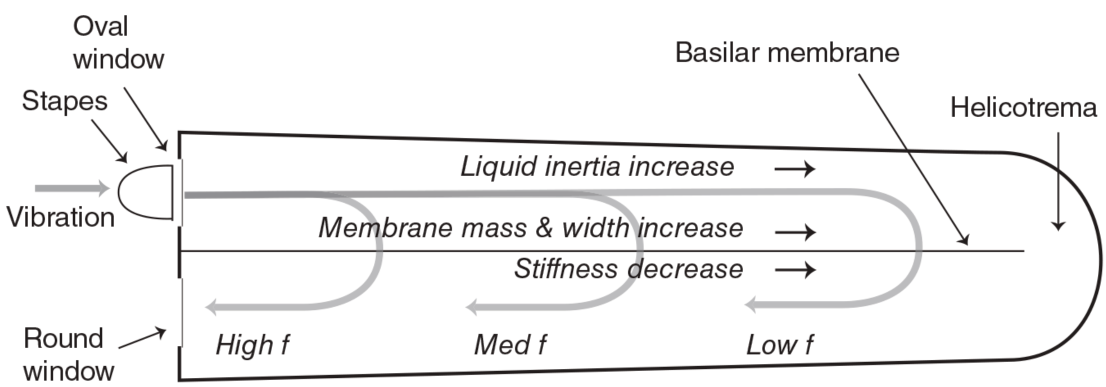


FIGURE 1.9: Displacement and frequency analysis on the basilar membrane.

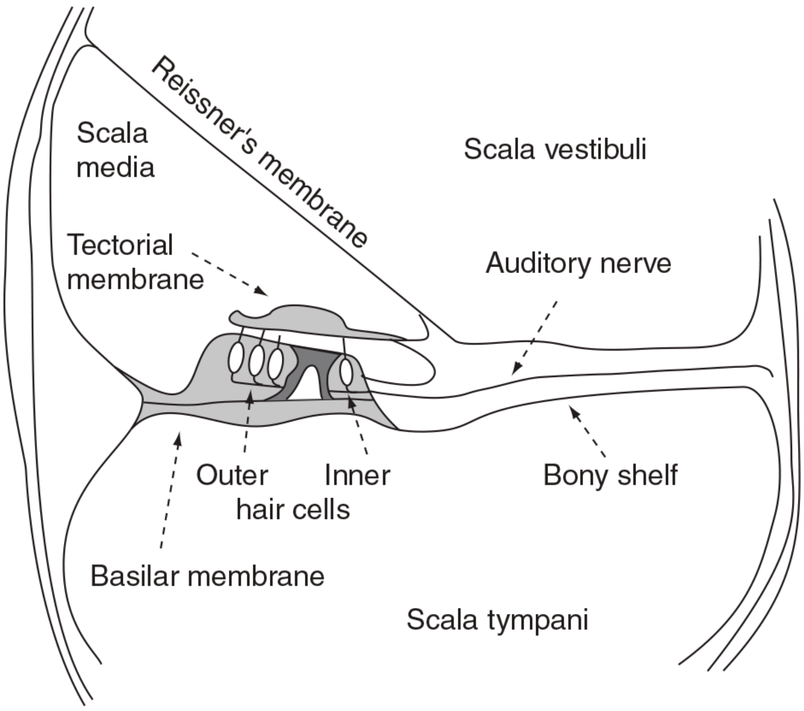


FIGURE 1.10: Cross-sectional structure of the cochlea.

Inner hairs cells lie in a single row along the length of the cochlea, while outer hairs cells lie in multiple rows toward the center of the basilar membrane, where vibration tends to be greater. The structure above the hair cells, known as the tectorial membrane, supports the stereocilia of the outer hair cells, which are embedded within it, but not the stereocilia of inner hair cells.

Inner hair cells trigger neural firing in response to vibrations. When the basilar membrane moves, stereocilia of the inner hair cells are bent, changing potentials and activating electro-chemical firings in the connected nerve. The vibration of the basilar membrane leads to the bending of the stereocilia which, in turn, causes the channels on the tops of the hair cells to open for K^+ ions [26]. The movement of such ions modulates the potential difference across the membrane of the cell, triggering the release of neurotransmitters at synaptic junctions between the inner hair cells and neurons of the auditory nerve.

The frequency-dependent behaviour of the basilar membrane means that it acts, together with the signal detection by the hair cells, as a mechanical band-pass filter. A vibratory pattern on the basilar membrane is transmitted by hair cells to the auditory nerve, thus coding frequency to place and then to neural fiber position.

The function of the outer hair cells is not completely known, but it is thought that it enables local amplification of the membrane vibration where it is being acoustically stimulated: this phenomenon is referred to as *cochlear amplifier*. This active role of the cochlea and hair cells is based on non-linear positive feedback: very weak signals are strongly emphasized, while, when the stimulus level increases, active boosting will decrease until the signal is no longer amplified [26].

A consequence of the active neural involvement in the cochlea is *otoacoustic emissions*. When a sound is presented to the ear, an echo can be recorded, and it can be delayed so much (more than 10 milliseconds) that it cannot be explained just by ear passive mechanics without involving neural activities. The resulting motion causes a faint sound, back-propagated out of the ear, called otoacoustic emission, and it is symptomatic of a healthy ear [6]. In many hearing impairments, the active role of the cochlea is damaged and such echo cannot be detected.

1.2.4 Perception theories

A series of nerve firings codes the information relative to the input sound; the organization of such information on its path to the brain plays a major role in subjective perception of pitch (property that allows a frequency-related ordering of sounds), timbre (perceived sound quality) and other sound features. To this moment, there is uncertainty regarding how the brain "decodes" the information. Several theories have been formulated trying to explain such process.

Place theory is related directly to the frequency analysis carried out by the basilar membrane. By this theory, the pitch of a sound is determined by the brain by looking at the places where the membrane vibrates due to resonance [23].

Temporal theory, instead, states that human pitch perception depends on the temporal patterns with which neurons respond to sound in the cochlea. This theory is based on the fact that the waveform of a sound with a strong pitch is periodic [9]. Therefore, the pitch of a pure tone is determined by the period of neuron firing patterns generated in the organ of Corti [23].

A joint place-temporal theory is nowadays considered [22]. Temporal theory well suits low-frequency situations, where the firing period is long enough to be phase-locked and decoded by the brain, while place theory can not explain perception below 50 Hz because pattern of vibration on membrane does not appear to change. As the frequency goes up (above 5 kHz), the period shortens, phase locking is lost and place theory becomes dominant.

It has been hypothesized that a probabilistic combination of both place and time cues is involved to decode the pitch of the incoming sound wave.

1.3 Frequency sensitivity range

The average standard frequency range of human hearing is reported in literature to be from 20 Hz to 20 kHz [29]. However, marked differences subsist between individuals and major changes are often introduced by the aging process.

In particular, the upper extreme of the range tends to reduce: it decreases (on average) to 16 kHz by the age of 20 and keeps reducing gradually from there [9]. This reduction is accompanied by a decline in hearing sensitivity at all frequencies with age, which affects low frequencies less than high frequencies. Hearing losses can also be induced by other factors, such as prolonged exposure to loud sounds (e.g. live concerts, discos, noisy environments).

Ear sensitivity to sounds of different frequencies varies over a vast sound pressure level (SPL) range, represented on a dB SPL scale:

$$\text{SPL} = 20 \log \frac{p}{p_{ref}} \text{ [dB]} \tag{1.1}$$

where p is the sound pressure level (in Pa) and p_{ref} is the reference pressure level (20 µPa), corresponding to the minimum sound level of a pure tone that an average human ear with normal hearing can perceive with no other sound present, at 1 atm and 25 °C.

The smallest amplitude of a tone causing an auditory event is called the hearing threshold (Fig. 1.11). This threshold curve is measured as the minimum audible field (MAF), defined as the "SPL of the weakest binaurally perceived sound from the front", and it varies with frequency: human ear is far more sensitive in the mid-frequencies range, where speech sounds lie, than at the high and low extremes.

1.3.1 Effective hearing area

A 0 dB value in the sound pressure level scale is set to be close to the hearing threshold for a 1-kHz pure tone. The sensitivity of hearing is best at about 3–4 kHz, and the hearing threshold increases below 250 Hz above 5 kHz [26]. The upper limit is about 130 dB, beyond which pain starts to be instantaneously felt by the listener and the hearing system is in danger of being injured, even for a short impulse of sound. Even at considerably lower levels (e.g. from 85 dB) long exposures result in hearing loss or damage.

Human speech, measured from one meter distance, is typically about 60–70 dB SPL, which is the optimal level for communication since it is not too loud to be harmful but is loud enough to produce a good signal-to-noise ratio (and to understand the conversation, as a consequence) in most environments. Such levels also provide good conditions for temporal and spectral analysis to the auditory system [26].

Although the full functional range of hearing is vast, just a portion of the total is used in common situations. Effective hearing areas for acoustic music and speech communication are later reported in Fig. 1.13.

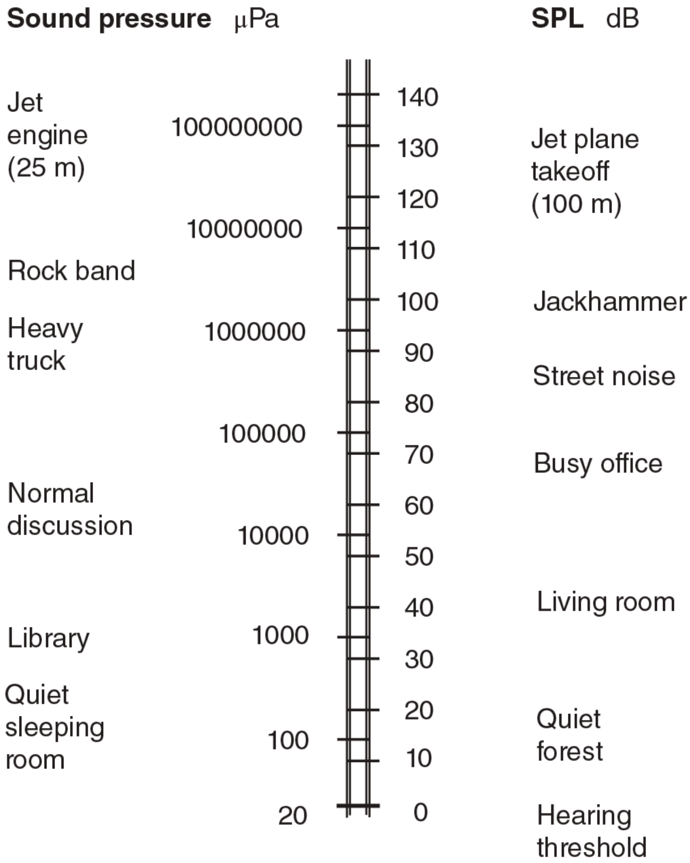


FIGURE 1.11: Sound pressure levels and related pressure values, from hearing threshold to threshold of pain, with examples of real world sounds.

1.4 Loudness perception

Loudness is the subjective perception of sound pressure. It is defined as the "attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud" [2].

It is a subjective measure and it is often confused with real measures such as sound pressure level or sound intensity. In fact, loudness is related to the sound pressure level, the frequency content and the duration of a sound [24], together with the health condition of the hearing system. Therefore, changes in the perception of loudness can be caused by multiple factors and may lead to different conditions, from bass enhancement to hearing loss. Different kinds of filters and weightings have been introduced, attempting to compensate real measurements to correspond to what is actually perceived.

The main units used to measure loudness are Sone (loudness level N) and Phon (loudness level L) [21]. Such scales are not SI units and are not proportional to each other. Sone is defined by stating that a loudness of 1 Sone is equivalent to the loudness of a 1-kHz tone at 40 dB SPL; doubling the perceived loudness doubles the Sone value [34]. Phon refers to the sound pressure level of a reference frequency of 1 kHz and matches dB levels of a similarly perceived pure tones along the spectrum. This implies that 0 Phon is the limit of perception and inaudible sounds have negative Phon levels [13].

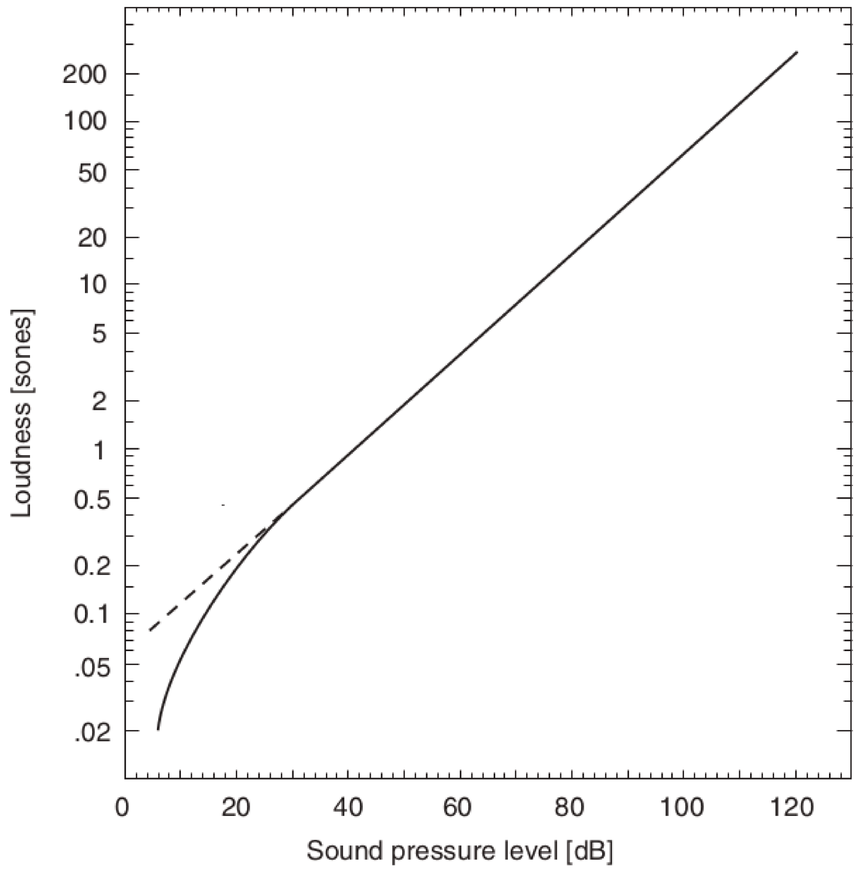


FIGURE 1.12: Loudness of a 1-kHz tone in sones as a function of the sound pressure level.

1.4.1 Equal-loudness level contours

Equal-loudness level contours (ELLC) provide measure of sound pressure levels (in dB), as the frequency changes, for which a human listener perceives a constant loudness (in Phon) when presented with pure tones. Two sine waves of differing frequencies have equal loudness level when perceived as equally loud by the average young person with no relevant hearing impairment.

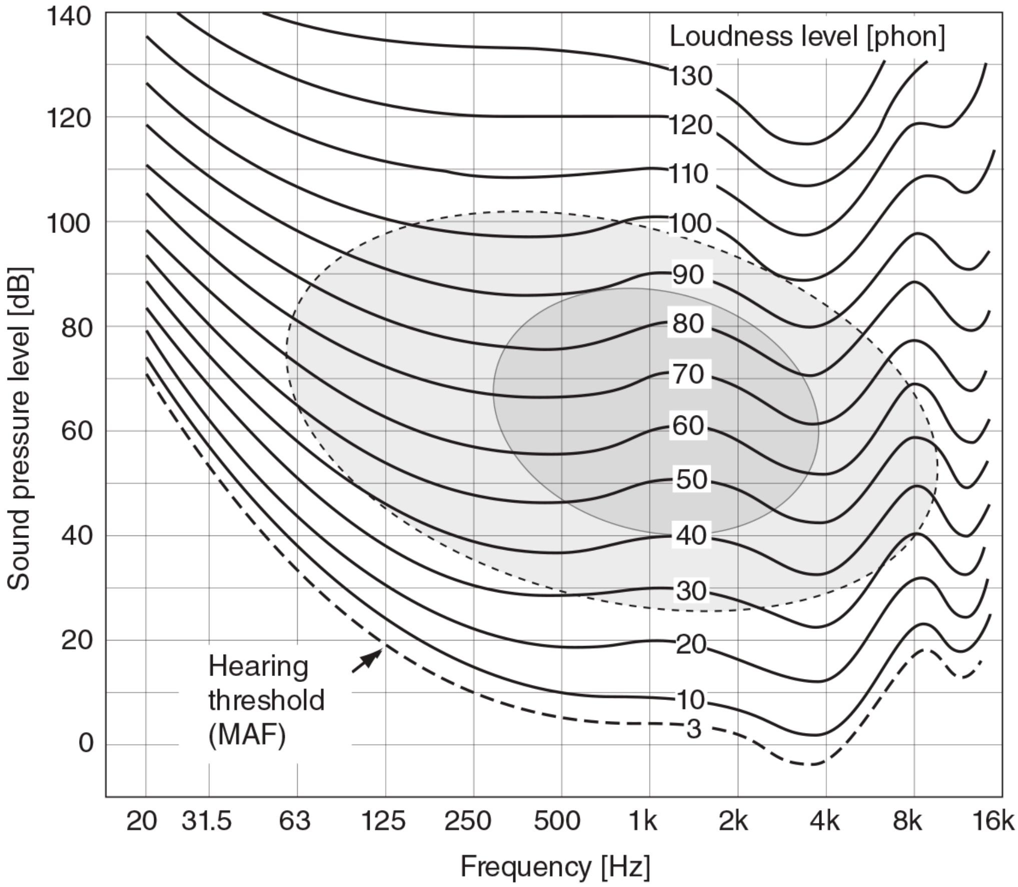


FIGURE 1.13: Equal-loudness contours, with typical acoustic music (light grey) and speech communication (dark grey) sound ranges.

The first research on equal-loudness level contours was conducted by Fletcher and Munson [5] in 1933; for this reason, ELLC are sometimes referred as "Fletcher-Munson curves", even though those studies have been surpassed and incorporated into newer standards. Fletcher and Munson first measured equal-loudness contours using headphones: test subjects listened to pure tones at various frequencies and over 10 dB increments in stimulus intensity. For each frequency and intensity, the listener also listened to a reference tone at 1 kHz.

The reference was adjusted until the listener perceived it at the same loudness as the test tone. The lowest equal-loudness contour (0 Phon) represents the quietest audible tone—the absolute threshold of hearing. In 1956 Robinson and Dadson [28] produced a new experimental determination which was believed to be more accurate and became the basis for an initial standard.

The most recent and definitive version of those curves is the one defined in the international standard ISO 226:2003 [12], which produced new curves by combining the results of studies by researchers in Japan, Germany, Denmark, United Kingdom and USA.

Equal-loudness contours are useful to show the sensitivity of the hearing system. The human ear proves to be most sensitive between 2 and 5 kHz, mainly because of the resonance of the auditory canal and displacements on the basilar membrane. It shows instead really low sensitivity in the low-frequencies range, which decreases further as the loudness level goes down.

1.4.2 Loudness compensation

Loudness compensation is the correction to be applied during playback of an audio source when it is reproduced at a different level from its recording and mastering level [25]. It is required to preserve the perceived spectral balance of sound, irrespective of playback volume level.

The need for loudness compensation arises due to the inherent non-linearity in the human hearing system. As the loudness level decreases, the low sensitivity of the ear to high and low frequencies may cause these signals to fall below hearing threshold. As a result, audio may become thin when played at low volumes, losing *bass* (low-frequency components) and *treble* (high-frequency components).

The loudness compensation feature that sometimes is offered by hi-fi audio equipment (Fig. 1.14) tries to apply a sort of counter-equalization, with arguable results so far [8].

FIGURE 1.14: Typical loudness compensation button available on audio equipment.

1.4.3 A-weighting

A-weighting is the most commonly used of a family of curves, defined in the standard IEC 61672:2013 [11], applied to instrument-measured sound pressure levels to account for the relative loudness perceived by the human ear. It arithmetically adds a table of values, listed by octave or third-octave bands, to the dB SPL measurement. The resulting octave band measurements are usually added with a logarithmic method to provide a single dBA (A-weighted decibel) value.

A-weighting slightly emphasizes the levels at mid frequencies and attenuates them at low and high frequencies. When the weighting curve is compared with the inverted equal-loudness curves, it is easy to see that the A weighting is just a very rough estimate, although providing numerous advantages in practical uses [26].

Other weighting curves are B, C, D and Z, even though they are less popular. In particular, Z-weighting is suitable for a sound produced with equal sound pressure across the whole frequency spectrum, while the C-weighting curve represents what is heard in the transient when sound is turned up (Fig. 1.15).

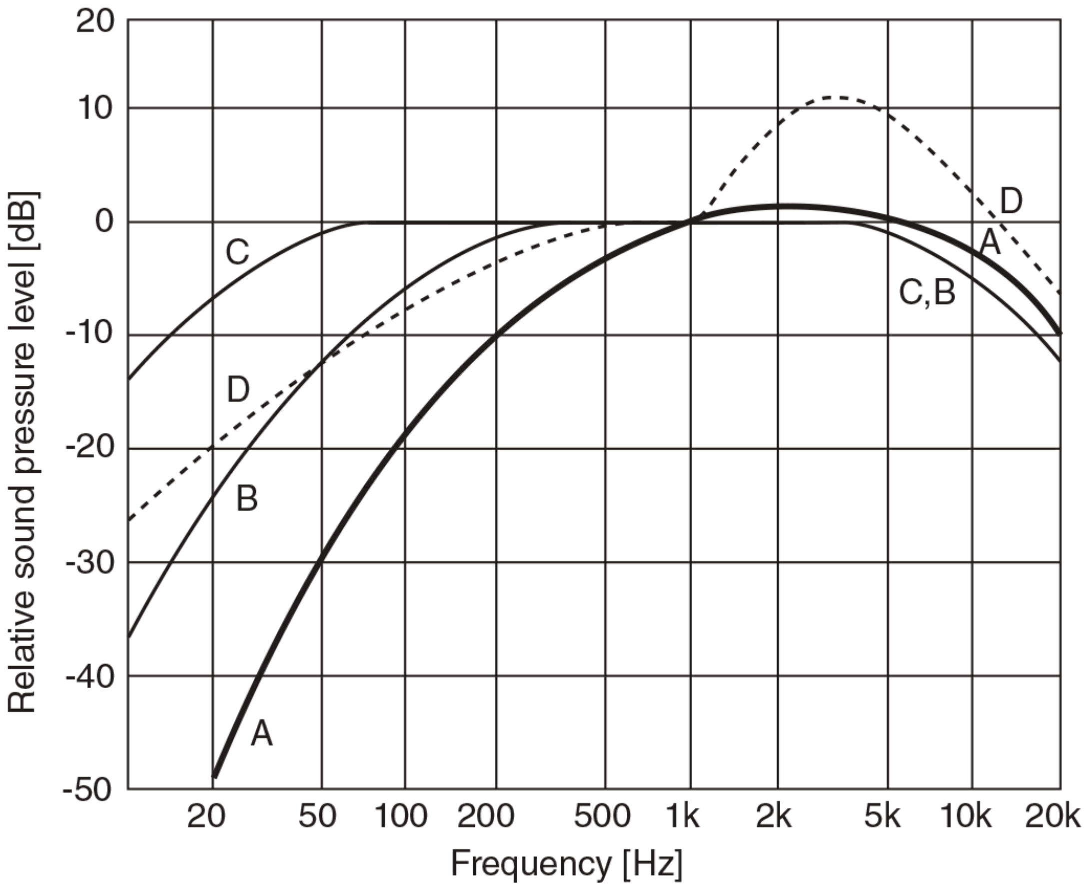


FIGURE 1.15: Relative levels of weighting curves A, B, C, and D for sound level measurement.

A-weighting is commonly used for the measurement of environmental or industrial noise, as well as when assessing potential hearing damage and other noise health effects at all sound levels. It is also used when measuring low-level noise in audio equipment, as alternative to ITU-R 468 noise weighting [15].

1.4.4 LKFS and loudness normalization

LKFS (loudness, K-weighted, relative to full scale) is a loudness measure designed to enable normalization of audio levels for broadcast delivery. It is described in ITU-R BS.1770-4 [14], which introduces a loudness gate based on a particular weighting, named "K", accounting for human perception.

ITU-R BS.1770-4 is being used (or planned to be used) for loudness measurement and normalization in broadcast, movie and home theaters and in music playback for streaming services like Spotify or YouTube. The goal of loudness normalization is to bring the average amplitude to a target level, countering the changes in loudness when listening to multiple songs in a playlist.

Depending on the dynamic range of the content and the target level, loudness normalization can result in peaks that exceed the recording medium limits. Dynamic range compression is often applied to prevent clipping, altering signal-to-noise ratio and relative dynamics.

Loudness normalization proves to be much more meaningful than classical peak normalization (which provides a dBFS measure), since it locates and takes into account in a different manner moments of silence or low-loudness during playback.

1.5 Listening tests

Being strictly related to a sensorial and perceptual sphere, psychoacoustics relies on listening tests to measure and analyze how sounds are perceived by humans [9]. That is because direct measurements are almost always impossible to be performed, due to ethical as well as practical reasons (extremely invasive surgery would be required). Also, higher-level processing made at cognitive level differs for every listener, so "objectiveness" is strongly questionable when discussing perception.

Subjective testing is typically a long and demanding process, requiring preparation and knowledge in order to provide valid results. A listening test should be performed to find preferences and identify problems of method or system and quantify response to a given stimuli [4]. It should not to be done when a similar test has been conducted or if there is an objective way of evaluation, like a direct measure.

1.5.1 Subjectiveness

When a psychoacoustic listening test is conducted, *subjectiveness* of the responses is a key word. Each answer corresponds to the opinion of a subject and all answers must be collected without judgment. Design of a listening experiment should be extremely careful, to ensure that the obtained results can truly reflect whatever aspect of the sound is under test. A behavioral response (i.e. a listening experience) is denoted a *dependent variable*; those aspects that might affect it are called *independent variables* [9].

All the independent variables must be controlled so that every observed effect can be attributed to some change in the specific independent variables under test. A schematic of the general working chain behind the design process of a listening test is reported in Fig. 1.16.

1.5.2 Design considerations

There are many aspects that should be considered during the design of a listening test.

- First of all, some characteristics of sound can affect the perception of the feature under test. For example, perceived pitch is affected by loudness, timbre and duration of the same audio sample.
- Particular care should be put in the choice of subjects regarding their expertise and their number, since it has a huge impact on the statistical significance of collected data. Also, some aspects like aging and hearing health must be taken under consideration.

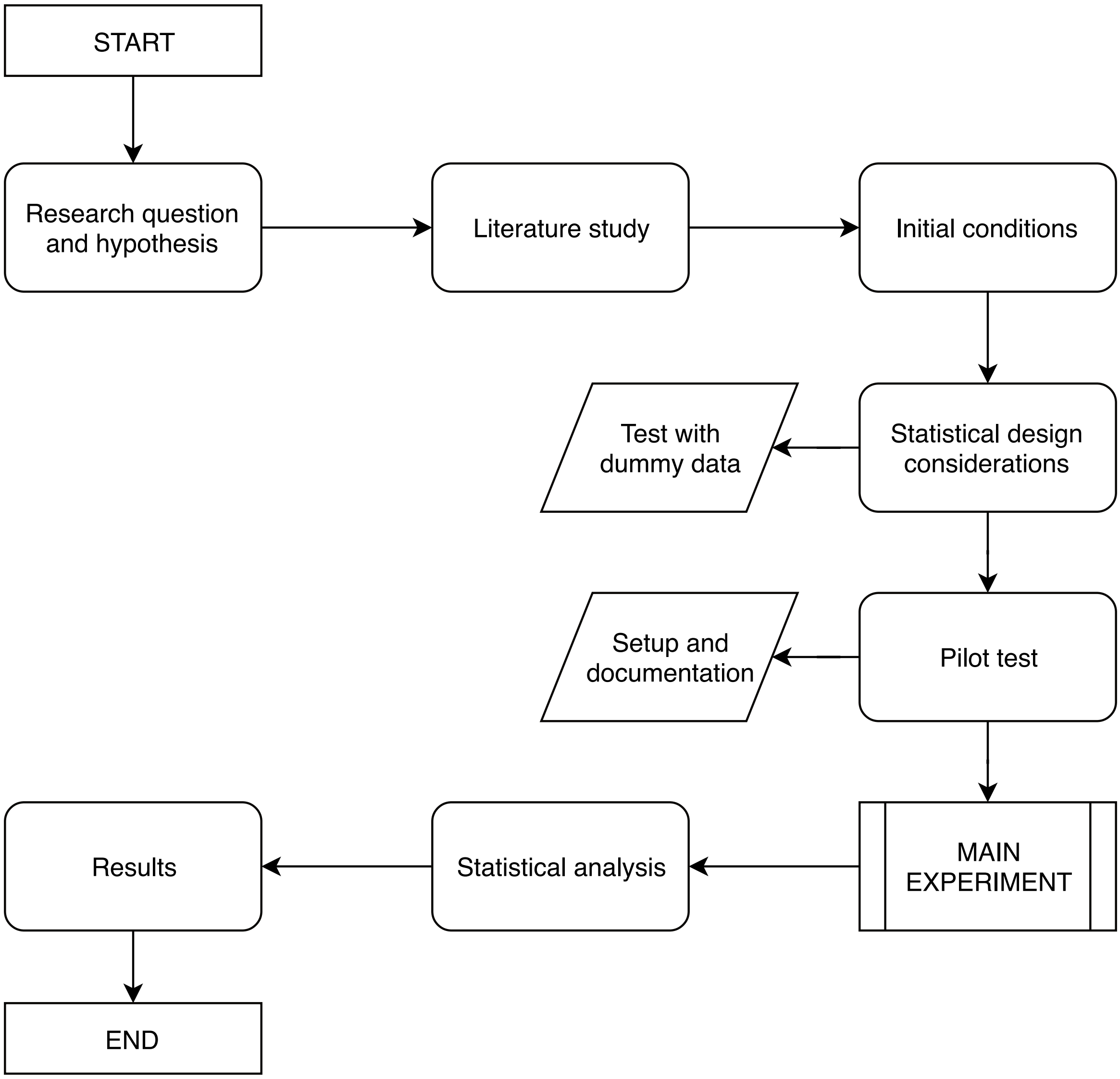


FIGURE 1.16: General steps in the design process of a listening test.

- Sound presentation is also relevant. The use of loudspeakers implies that the room acoustics will affect sound waves arriving at each ear, while the use of headphones avoid this concern. Background acoustic noise in the listening room, even localized, can have a huge impact on the results.
- Subjects can become tired or distracted after some time [4], so test duration is also a key element.

It is also obvious that sound samples for a listening test must have high and consistent quality and that the test environment should allow the subjects to feel comfortable.

All potential independent variables must be identified and put under control prior to the test. Commonly, a pilot test is run with a small number of listeners to check the procedure and the presence of any additional independent variables [4].

1.5.3 Test procedure

In order to conduct a listening test with success, i.e. to obtain meaningful results, some basic rules should be observed.

Before the actual listening test, the subjects must be instructed, giving them all necessary information and explanations about the planned experiment. The purpose should be explained only if such information will not influence their judgment. Only if the task is understood, the subjects will feel confident during the test and be capable of completing such a task in a reliable manner. The duration of the test is also an important thing to communicate, in this sense.

After the briefing, a training of the subjects can be conducted, depending on the difficulty of the task and the experience of the subjects. In a training session, samples can be presented to the subjects in advance, to prepare for the actual test and know what to expect. The training should not be too long in order to avoid a loss of concentration or cause boredom which will affect the results.

The test supervisor must make sure that the subjects are not disturbed during the test (a break is commonly allowed to subjects) and must be available for inquiries during the test.

After the test, feedbacks on the test can be collected from subjects, in order to improve the test procedure in future deployments. Of course, the privacy of all subjects must be fully respected before, during and after the test.

1.5.4 Screening and statistical analysis

A wide range of statistic methods is available to evaluate results from a listening test and allow the gathered data to be summarized and presented in a concise and easy-readable form.

Generally, all test methods require the subjective judgments to be translated into numeric values, with data from inconsistent subjects being discarded. The screening process should be extremely cautions: discarding too many subjects can introduce a bias in the results and, as a consequence, cause the invalidation of the test.

Obtained numerical data can be represented and compared graphically, giving a general overview of the results and helping the supervisors in their understanding. Different statistical operators can be introduced: average, median, variance, quartiles, etc. Any operation should be carefully considered before being applied and be well documented so it is possible to reproduce the results and how they were obtained in a second moment [4].

Following this process, a meaningful interpretation of the results is possible.

Chapter 2

Loudness compensation function

In the previous chapter, loudness perception has been presented and discussed, together with the need for compensation to preserve the balance of the different perceived frequency components. In this chapter, a new method for adaptive compensation of the low frequencies during sound reproduction is shown, starting from the equal-loudness contours in order to obtain an ideal compensation function.

2.1 Equal-loudness contours interpolation

Standard ISO 226:2003 is used as reference for this method. It specifies the sound pressure levels of a pure tone, in function of frequency, perceived as equally loud by human listeners in free space [12], and provides a polynomial function for the contours:

$$L_p = \frac{10}{a_f} \log_{10} A_f - L_u + 94 \tag{2.1}$$

$$A_f = 4.47 \cdot 10^{-3} \cdot [10^{\,0.025 L_n} - 1.15] + [0.4 \cdot 10^{\,0.1(T_f + L_u) - 9}]^{\,a_f}$$

where:

- L_p is the sound pressure level (in dB) of a pure tone;
- L_n is the loudness level (in Phon);

- f is the frequency of the pure tone;
- T_f is the hearing threshold (in dB);
- a_f is the exponential factor, accounting for loudness perception;
- L_u is the magnitude (in dB) of the frequency response, normalized at 1 kHz.

The data range provided by the standard is 20 to 90 Phon, in a frequency span going from 20 Hz to 12.5 kHz, as shown in Fig. 2.1.

From the contours, it is easy to see how the sensitivity of human perception changes non-linearly with frequency and how the bass range is the most heavily affected part of the spectrum. Data from ISO 226:2003 is linearly interpolated with 1-Phon steps to provide intermediate curves (Fig. 2.1). Interpolation is run in between the 20-120 Phon extremes, since data below and above such values are not meaningful for the purpose of this work. A secondary spline interpolation can be done in order to obtain more frequency points, but it is not required yet.

2.2 Sensitivity function

The main idea behind the proposed compensation method is quite straightforward. A *sensitivity function*, relative to the playback level, can be derived from the equal-loudness contours, providing a more clear estimation of how much perception varies as frequency and listening level are changing. An inverse curve can then be found and used as a trace-guide for the design of a digital filter, capable of modifying the perceived sound and restore the original spectral balance.

Starting from the interpolated equal-loudness contours, it is possible to normalize each curve with respect to SPL at 1 kHz in order to evaluate the sensitivity of human hearing for different loudness levels, i.e. to obtain a sensitivity function S:

$$S\ (f, L_n) = -L_p(f, L_n) + L_p(1000, L_n). \tag{2.2}$$

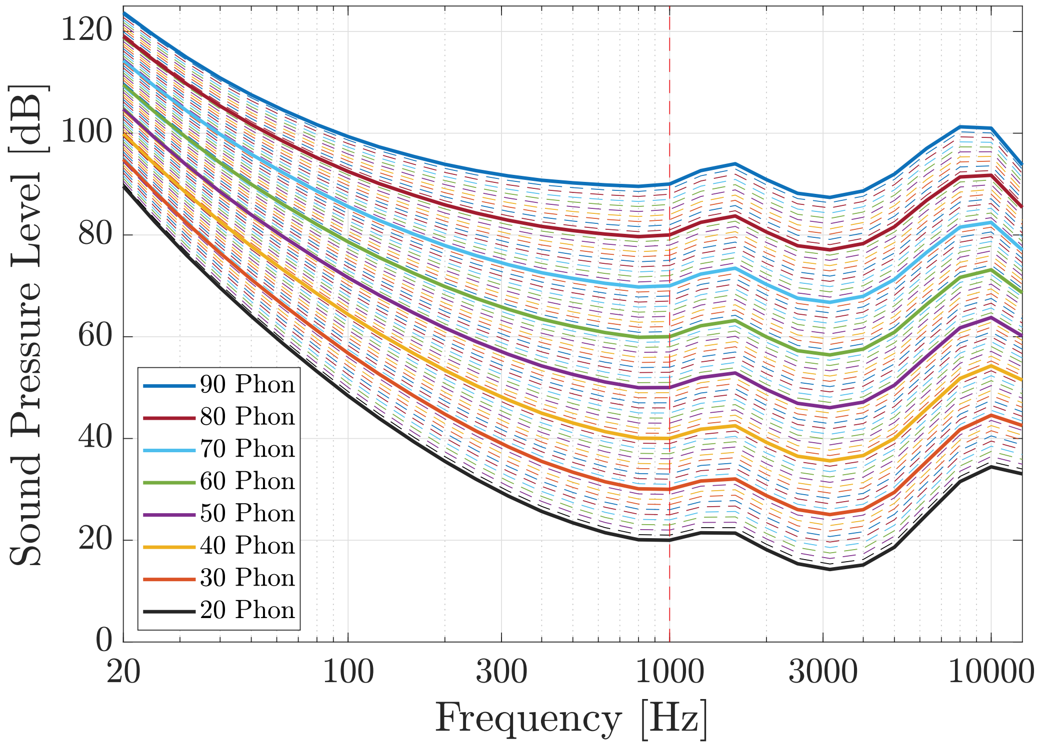


FIGURE 2.1: Standard (thick lines) and interpolated (dashed lines) equal-loudness contours for 20–90 Phon range from 20 Hz to 12.5 kHz.

Curves obtained from such a function are shown in Fig. 2.2. The difference in perception with respect to sound pressure level at 1 kHz corresponds to the gain (or attenuation) to be introduced in order to have a flat response in loudness perception.

It is well known (and observable in Fig. 2.2) how the bass and sub-bass ranges are much more affected by perception change than mid and high frequencies, especially when the sound reproduction level goes down. For example, the maximum deviation from 1-kHz value is around 15 dB for high frequencies, while it can drop down to 70 dB for low frequencies. Moreover, the difference between 80 Phon curve and 20 Phon curve at 50 Hz is approximately 20 dB, while the difference between the same curves at 3 kHz is less than 5 dB.

Consequently, it is easy to understand the importance of finding a way to adapt the compensation to the listening level of the playback audio track. An over-boosting

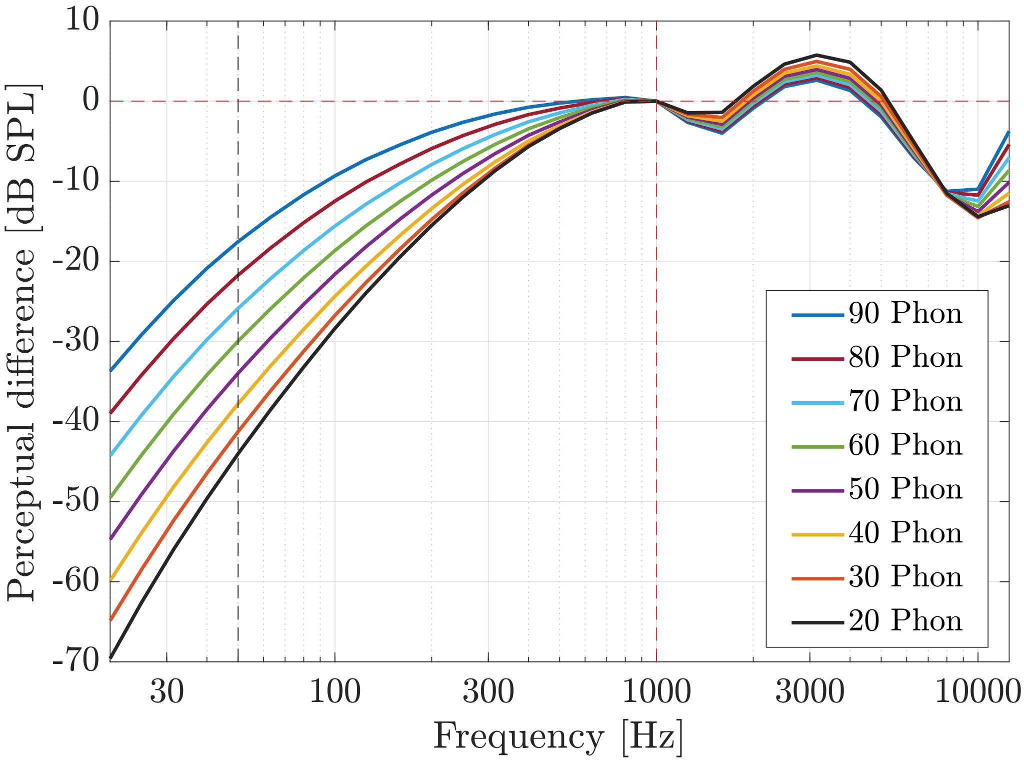


FIGURE 2.2: Sensitivity function derived from ELLC, showing perceptual difference as frequency changes for different loudness levels.

might be pleasant at low levels, but it must be kept in mind that the goal is fidelity; an under-compensation is obviously an even worst situation.

2.3 Compensation function

As said before, the mastering level L_M and the listening level L_L for music are usually different, the first being louder than the latter [16]. The goal is then to compensate the perceived spectral balance at L_L in such a way that it matches the perceived spectral balance at L_M.

A relationship between the two levels can be derived starting from S: a correct compensation should be applied to L_L sensitivity curve in order for it to match the L_M sensitivity curve. For each L_M-L_L pair, it is possible to identify a *difference curve*

ΔL_p representing the difference in perception between the two levels:

$$\Delta L_p(f, L_M, L_L) = L_p(f, L_M) - L_p(f, L_L) - N_f, \tag{2.3}$$

where N_f is the normalization factor for the sensitivity function, according to (2.2):

$$N_f = L_p(1000, L_M) - L_p(1000, L_L) = L_{M_{dB}} - L_{L_{dB}}. \tag{2.4}$$

Inverting (2.3), a balancing curve is finally obtained. It corresponds to the magnitude response of an ideal filter that perfectly balances the perception of spectral components as intended by the mastering process:

$$\begin{aligned} H(f, L_M, L_L) &= -\Delta L_p(f, L_M, L_L) \\ &= -L_p(f, L_M) + L_p(f, L_L) + N_f \\ &= -L_p(f, L_M) + L_p(f, L_L) + L_{M_{dB}} - L_{L_{dB}}. \end{aligned} \tag{2.5}$$

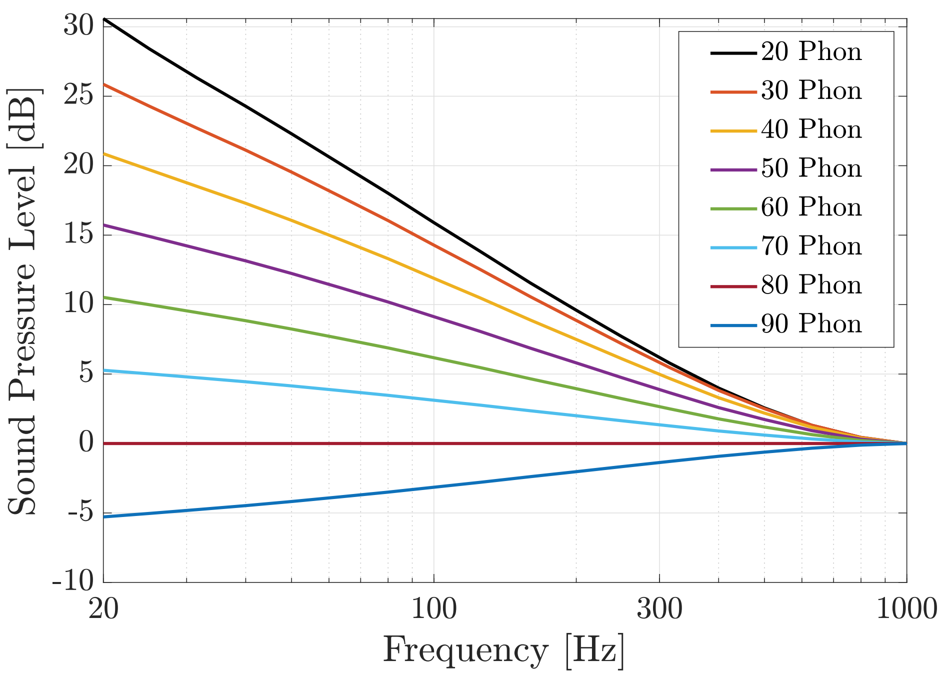


FIGURE 2.3: Derived trace-guides at $f \leq 1$ kHz, $L_M = 80$ dB SPL.

Notice how the slope of the curves in Fig. 2.3 is always lower than 20 dB/decade and, in general, lower than 10 dB/octave and how bass reduction, instead of boost, is required for $L_M < L_L$. This is significant for situations where music can be played at high levels, e.g. live concerts, discos or noisy work environments.

A MATLAB implementation for the equal-loudness contours provided by the standard, for their interpolated version and for the derived compensation function can be found in Appendix A.

Chapter 3

Filter design

Since a set of curves (a *trace-guide*) has been obtained in the previous chapter, the next step on the line is the actual filter design. The goal is to identify a type of digital filter whose magnitude response resembles the trace-guide and that can easily adapt to variations in the reproduction level. Such a filter should be found with the lowest possible order and complexity, in order to have minimum impact on the playback system and allow a real-time implementation.

In this chapter, different kinds of filters are evaluated for the task. Parameter optimization is also run under particular conditions to improve the performance of the chosen filter.

3.1 Filter specifications

Given a set of specifications, the first step towards the desired filter is decide between either an IIR or a FIR implementation. After that, a subset of candidate filters can be considered, working and altering their parameters to obtain the target curve.

3.1.1 IIR filters

Infinite Impulse Response (IIR) filters are the most obvious choice in term of efficiency for audio DSP applications [37]. Processing typically is low-demanding in terms of memory and operations, allowing the implementation of such filters in low-cost, low-power architectures and products hitting the markets. Anyway, they require floating-point architecture to avoid that the approximation of filter coefficients causes major changes to filter behaviour, in particular instability.

3.1.2 FIR filters

Finite Impulse Response (FIR) filters enable linear-phase filtering and support integer math, but require an higher delay and and more memory on the DSP interface. Furthermore, FIR filters can be limited in resolution when working with low frequencies, influencing the quality of the filter coefficients [37].

For those reasons, FIR filters are not considered for the implementation of the proposed loudness compensation method, and in general are to be avoided for the majority of audio DSP applications.

3.1.3 Digital filter derivation

Digital filters can be derived:

- from known analog filters, converting a transfer function with analog poles/zeroes from the Laplace-domain to the z-domain and obtaining a difference equation using one of the common transformations:
  - bilinear transform;
  - MZT/MPZ (*Matched Z-transform/Pole-Zero Mapping*) method [32] [33];
  - impulse invariance method.
- by choosing appropriate locations for the poles and zeroes on the unit circle in the z-domain, thus imposing desired corner frequencies and slope [31].

IIR filters are typically implemented in Direct Form I (Fig. 3.1), as this gives the best Signal-to-Noise Ratio (SNR).

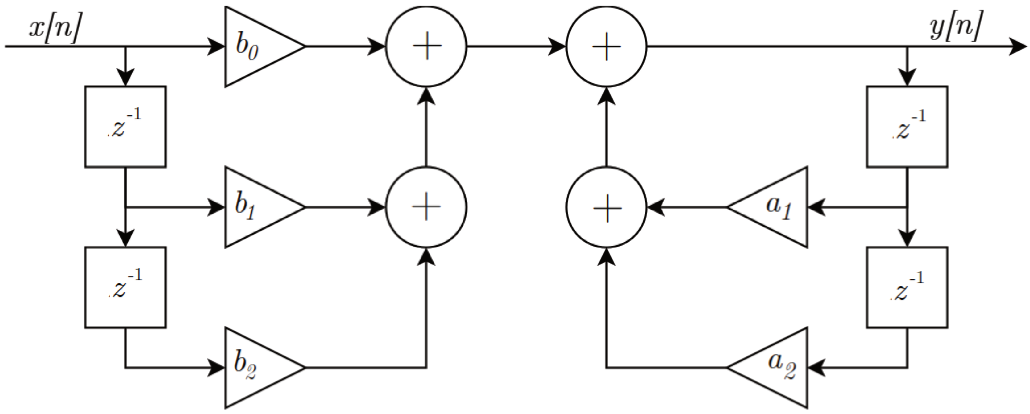


FIGURE 3.1: Direct Form I implementation of a second order digital filter.

3.2 Fractional order filters

From Fig. 2.3, it can be observed that the target magnitude response slope is less than 20 dB per decade, for every curve. This suggests to look for filters with an order smaller than one – the so called *fractional order filters* – and with a low-pass behaviour.

Fractional order filters (FOF) are neither well documented nor well described in DSP literature at the moment of writing [36], but a design process for digital fractional order filters has been proposed by Nielsen [19] and a MATLAB implementation is reported in Appendix B. An analog arbitrary-order low-pass transfer function for fractional order filters can be obtained as:

$$H(s) = \frac{s+\omega_{z_1}}{s+\omega_{p_1}} \cdot \frac{s+\omega_{z_2}}{s+\omega_{p_2}} \cdot \frac{s+\omega_{z_3}}{s+\omega_{p_3}} \cdot ... \tag{3.1}$$

under the following conditions:

- $G_0/G_\infty = \omega_{z1}/\omega_{p1}$;
- $G_\infty = 1$;
- $\omega_{p_i} < \omega_{z_i}$.

The first condition defines the slope of the response; the second ensures that mid and high frequencies are not affected by the filtering; the third is required for the desired shape. A following bilinear transformation (or one of the other method described above) allows to obtain the digital filter transfer function $H(z)$.

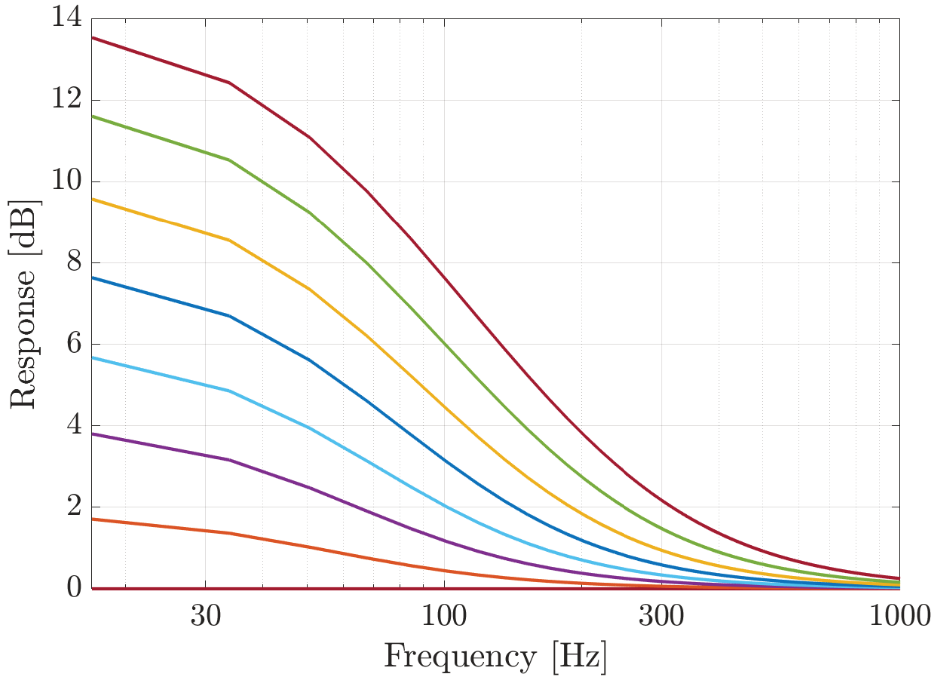
Different transfer functions obtained from a single pole-zero couple, describing what can be called a first "order" FOF, are reported in Fig. 3.2. Parameters determining the filter behaviour are the gain G_0 and the frequency of the pole p_1 (the corresponding zero z_1 is computed from p_1, according to the conditions stated above). G_0 impacts on the slope of the response (Fig. 3.2a), while the position of the pole mainly affects the cut-off frequency of the filter (Fig. 3.2b).

Anyway, a single pole-zero couple produces curves whose shapes can not match the trace-guide. Slopes can be gradually reduced by introducing multiple pole-zero couples with sharper corner frequencies, choosing poles and gains properly. Second and third "order" FOFs, i.e. obtained from respectively two and three pole-zero couples, are the optimal choices in terms of trade-off between response shape and computational complexity.

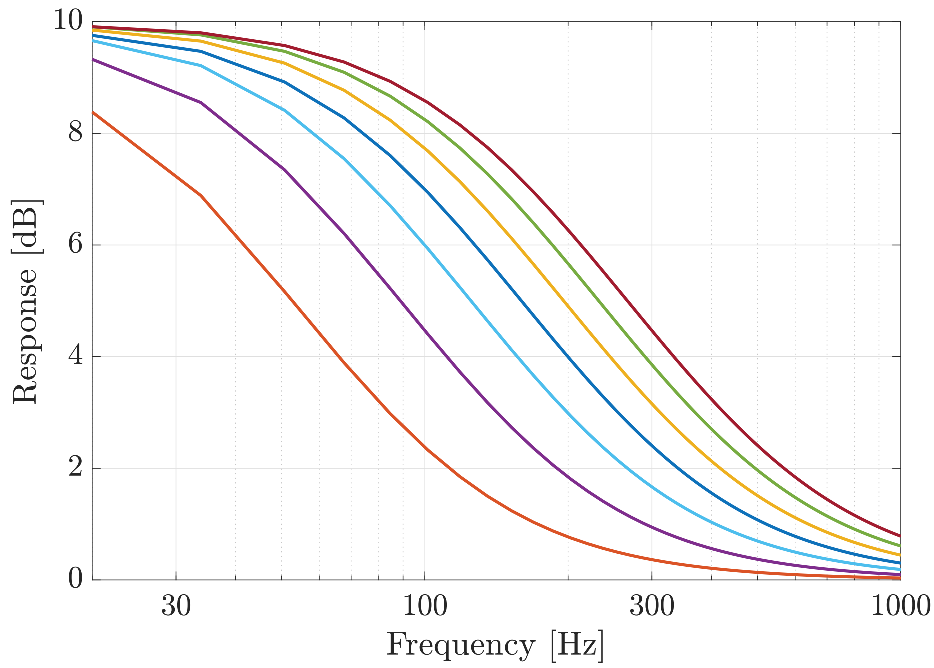
After some trial-and-error steps, the frequency response curves from chosen pole-zero couples seem to already fit quite decently the equal-loudness contours trace-guides (Fig. 3.3, 3.4). A small error is introduced towards 1 kHz, while the lowest frequencies are under-compensated. This last effect is actually preferable, because in such spectrum range there is mainly noise and hardly anything else interesting to be heard.

It is worth mentioning that the transfer function shown in (3.1) gives a n-th "order" FOF from a cascade of first "order" FOFs.

Although providing similar results as shelving filters (see next section), the latter should eventually be preferred, due to the wide-spread use and lower complexity (always a good thing when discussing real-time audio applications). Future studies might provide interesting results, also relevant for this work.



(A) Single pole/zero couple, fixed pole ($f = 25$ Hz) and variable gain.



(B) Single pole/zero couple, fixed gain ($G_0 = 10$ dB) and variable pole.

FIGURE 3.2: Effects of gain and pole frequency on the response of a fractional order filter.

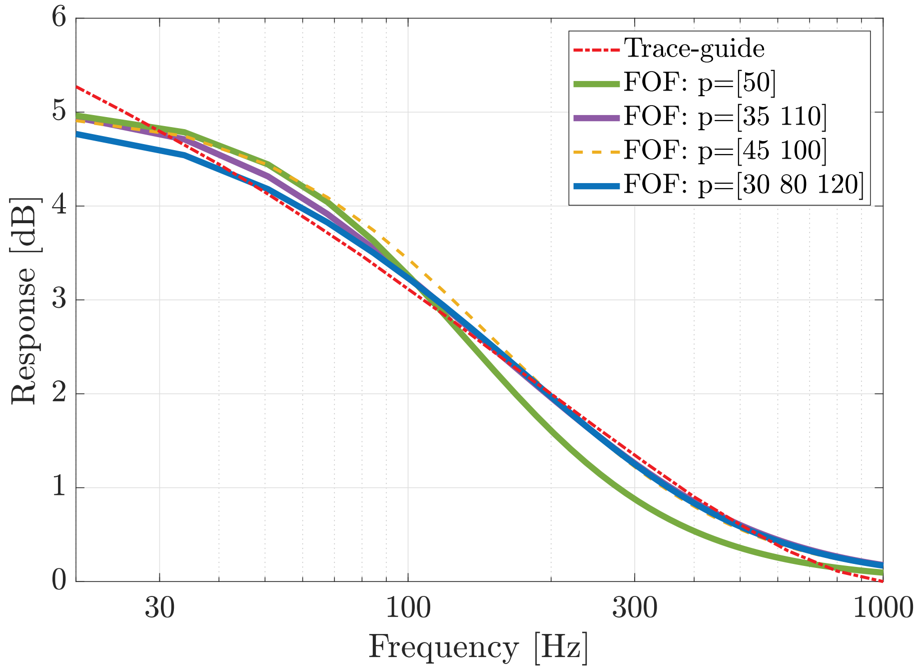


FIGURE 3.3: Comparison between trace-guide from ELLC and some computed FOF responses, ML = 80 dB SPL, LL = 70 dB SPL

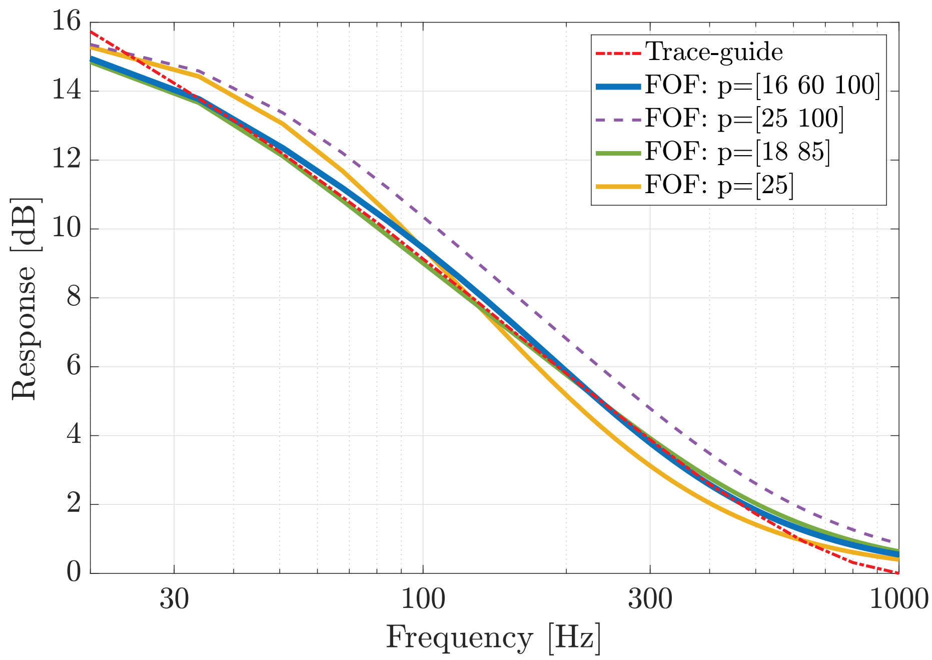


FIGURE 3.4: Comparison between filter trace from ELLC and some computed FOF responses, ML = 80 dB SPL, LL = 50 dB SPL

3.3 Shelving filters

A shelving filter boosts or attenuates the magnitude of an input signal in a certain frequency band - either the lowest-frequency band or the highest-frequency band - without necessarily cutting out the harmonics in that band as a typical low-pass/high-pass filter would do. Depending whether it works on the the bass range or the *treble* (high frequency), it is referred as either, respectively, *low-shelving* or *high-shelving* filter.

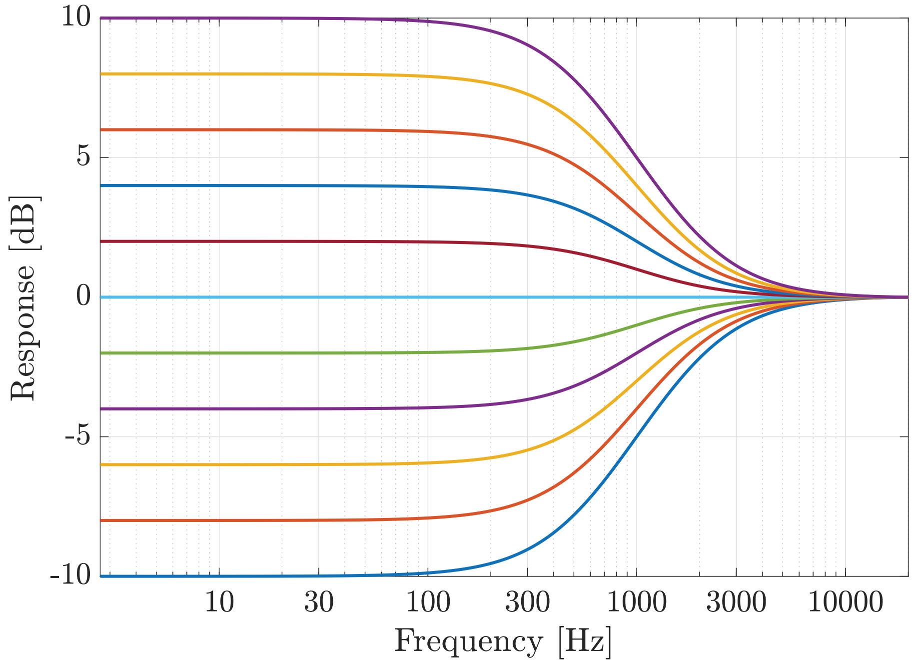
This type of filters is largely used in parametric equalizers, due to the smooth transition of the filter response between affected and unaffected regions and the simple implementation. A classic parametric EQ presents two knobs to the user, one for bass and one for treble[1], through which it is possible to alter the filter response and thus its effect on the sound reproduction.

Simplicity comes from the fact that the behavior of a shelving filter is completely described by just the gain G and the crossover frequency f_c (often also called *corner* or *cut-off* frequency), as can be seen from Fig. 3.5. The gain parameter affects the response at low frequencies and the slope (Fig. 3.5a), while the crossover frequency parameter relates to the width of the response, i.e. its frequency span (Fig. 3.5b).

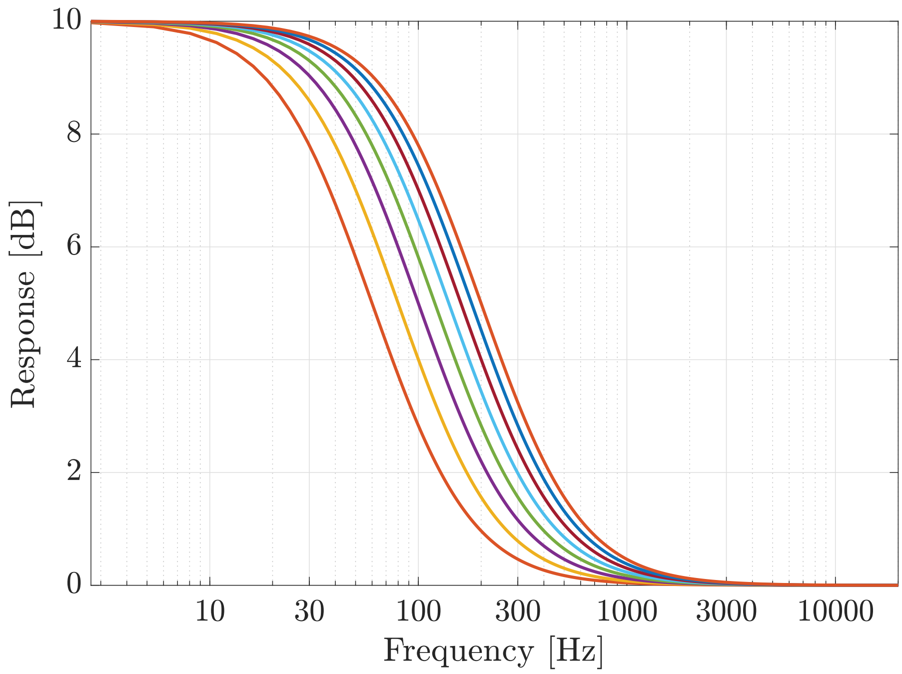
Transfer functions for both first and second-order low-shelving digital filters have been derived by Välimäki and Reiss [37] and a MATLAB implementation for both is reported in Appendix B. They can be written as:

$$\begin{aligned} H_{LS,1}(z) &= \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1}} \\ &= \left[\frac{G\Omega + \sqrt{G}}{\Omega + \sqrt{G}} + \frac{G\Omega - \sqrt{G}}{\Omega + \sqrt{G}} z^{-1} \right] \left[1 + \frac{\Omega - \sqrt{G}}{\Omega + \sqrt{G}}\, z^{-1} \right]^{-1}, \end{aligned} \tag{3.2}$$

[1]Typically, one or more knobs adding peaks/notches in the mid-frequencies range are also presented in the majority of audio equipment.



(A) Variable G, constant $\omega_c = 1$ kHz



(B) Variable ω_c, constant $G = 10\,\text{dB}$

FIGURE 3.5: Effects of gain and crossover frequency on the response of a first order low-shelving filter.

$$H_{LS,2}(z) = \frac{a_0\, b_0 + a_0\, b_1 z^{-1}}{a_0 + a_0\, a_1 z^{-1}}$$
$$= \sqrt{G}\frac{\sqrt{G}\Omega^2 + \sqrt{2G}\Omega + 1 + 2(\sqrt{G}\Omega^2 - 1)z^{-1} + (\sqrt{G}\Omega^2 - \sqrt{2G}\Omega + 1)z^{-2}}{\sqrt{G} + \sqrt{2G}\Omega + \Omega^2 + 2(\Omega^2 - \sqrt{G})z^{-1} + (\sqrt{G} - \sqrt{2G}\Omega + \Omega^2)z^{-2}}, \tag{3.3}$$

where $\Omega = \tan(\omega_c/2)$ and $\omega_c = 2\pi f_c/f_s$.

An arbitrary choice of gain and pole allows to verify the goodness of this type of filter; a comparison of different filter responses for a chosen listening level can be found in Fig. 3.6 and 3.7. The trace-guide is interpolated with a spline function to provide more frequency points. As it can be seen:

- the first-order shelving filter presents a fairly good approximation of the trace-guide, with small over-compensation in the sub-bass range and a slightly more relevant under-compensation in the bass-to-mid range;
- an even better result is achieved with a cascade of two first-order filters, which is able to follow the trace-guide along all the considered interval;
- second-order low-shelving returns curves which are too steep to approximate the trace-guide.

So forth, second-order shelving filter is not considered. A small error is introduced towards the mid frequencies, but (as said in the previous section) it is easily correctable in other manners, like a peak/notch filter in a parametric equalizer.

Moreover, the flat response towards the lowest frequencies avoids saturation and SNR degradation via noise over-amplification, which would have been introduced following the trace-guide in that part of the spectrum.

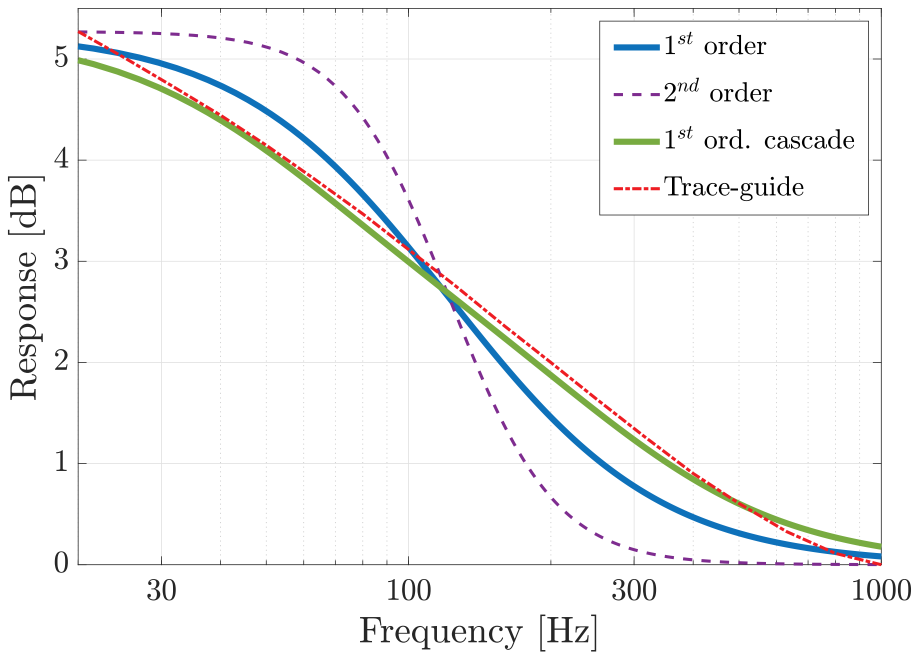


FIGURE 3.6: Comparison between trace-guide and shelving filters, $L_M = 80\,\text{dB SPL}$, $L_L = 70\,\text{dB SPL}$.

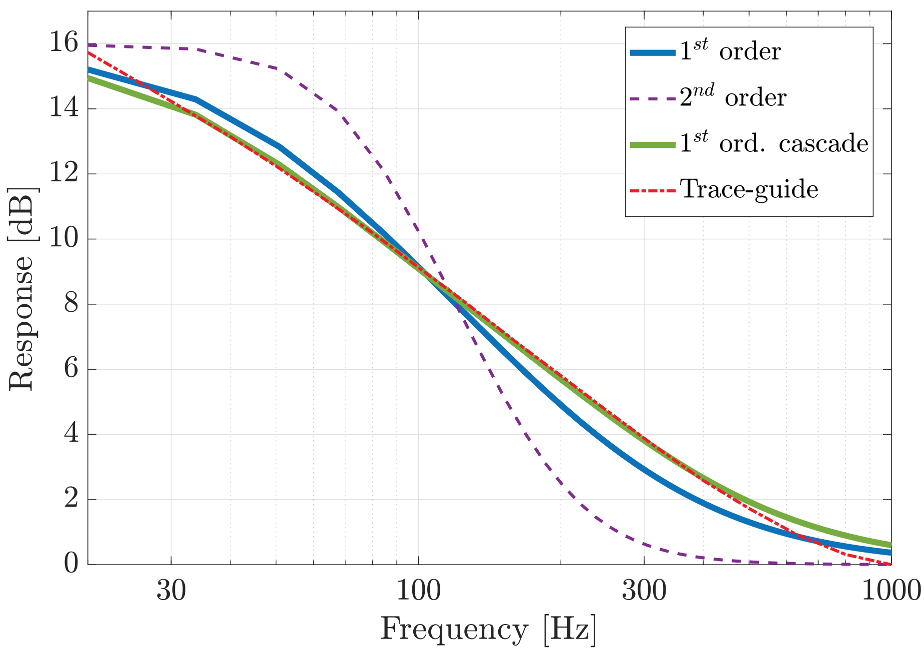


FIGURE 3.7: Comparison between trace-guide and shelving filters, $L_M = 80\,\text{dB SPL}$, $L_L = 50\,\text{dB SPL}$.

3.4 Optimization of filter parameters

Now that the filter type is chosen, the optimal filter parameters should be found (G, ω_c) in order to produce the best shapes with minimum deviation from the trace-guide. In case of a cascade of two first-order shelves, there are four parameters: (G_1, G_2, ω_{c_1}, ω_{c_2}). However, due to the goal of the compensation method, some things might be taken into account in order to simplify the optimization problem:

- mastering levels typically lie in a very short interval (80–85 dB SPL), so difference curves used as trace-guides will not differ much for levels in such a range;
- it is reasonable to choose trace-guide sound pressure level at 20 Hz as filter G; in case of a filter cascade, the product of the gains (the sum, in the logarithmic domain) should match such a value.

$$G_{\mathrm{dB}} \left(= \sum_k G_{\mathrm{dB}_k} \right) = \Delta L_p(20, L_M, L_L). \tag{3.4}$$

3.4.1 Optimization algorithm

The choice and the description of the optimization algorithm is not a critical task for this method, but it is worth of a digression.

No major constraints are put on the accuracy or the timing, but a good pick would be an algorithm providing high-quality solutions generated in a reasonable time and that has some previous implementation in filter design and parameter optimization literature.

The use of genetic algorithms (GA), already applied in many subjects such as pattern recognition, robotics, biology, and medicine, has been suggested for this task by Ahmad [1]. GAs are used in optimization and search problems and are members of the larger group of evolutionary algorithms. In fact, they are based on the process of natural selection, using operators such as mutation, termination, selection and crossover.

The process of natural selection starts by targeting the fittest individuals from a population. They produce offspring, which inherit the characteristics of the parents and will be added to the next generation. The better the parents' fitness, the higher is the chance for the offspring to survive. This process keeps on iterating and at the end, a generation with the fittest individuals is obtained. Five phases are considered before termination: initial population; fitness; selection; crossover; mutation (Fig. 3.8).

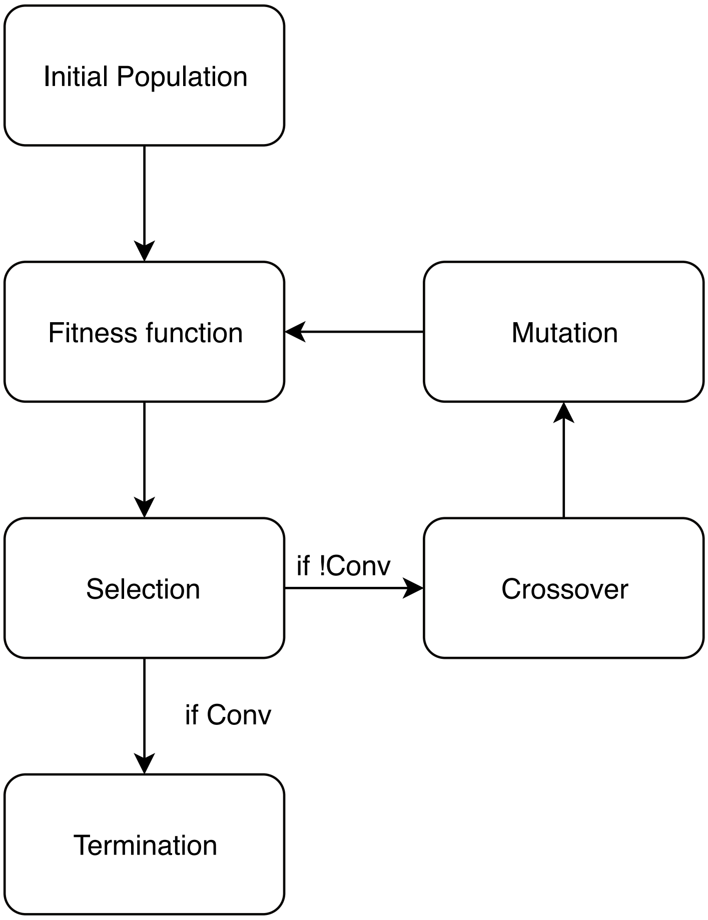


FIGURE 3.8: Schematic diagram, reporting the main phases of the genetic algorithm.

1. A set of individuals or *phenotypes* (a set of solutions for a problem), which is called *population*, is chosen to start. A phenotype is characterized by a set of parameters or *genes*, which are joined into a string to form a *chromosome* (a solution).

2. A fitness function determines how fit ("how good it is to compete with others" [1]) a phenotype is, assigning a fitness score to it which is used to determine the probability of that phenotype to be selected for reproduction.

3. The fittest phenotypes (*parents*) are selected and their genes are passed onto the next generation.

4. Offspring are created by exchanging the genes of parents among themselves until a crossover point, chosen randomly in between parents' genes, is reached and then they are added to the population.

5. Some of the genes of the new offspring can suffer a low-chance mutation, which has to occur to keep variance within the population and prevent premature (wrong) convergence.

6. Termination arrives if the population does not produce significantly different offspring from the previous generation, i.e. has converged.

GAs are based on Darwin's principle of natural selection, which is a slow process per se and, as a result, these algorithms often need a larger amount of computations. However, they offer interesting advantages over classical optimization algorithms, such as gradient-based and Newton-type ones.

In particular, GAs are capable of detect local sub-optimal solutions that can be discard in favor of more promising local solutions; therefore, they are more likely to obtain better solutions in multimodal problems. By contrast, classical optimization algorithms, although being very efficient, are not capable of discarding inferior local solutions in favour of more optimal ones.

A practical variant of the general process of constructing a new population is to allow the best phenotypes from the current generation to carry over to the next, unaltered. This strategy is known as *elitist selection* and guarantees that the solution quality obtained by the GA will not decrease from one generation to the next.

3.4.2 Crossover frequencies

Starting GA runs over the 80–85 dB SPL interval of mastering levels show that the optimal solution for gain balance in the shelving cascade is really close to an equal weighting. So it is safe to assume, in first approximation:

$$G_1 = G_2 = \frac{1}{2}\,G. \tag{3.5}$$

In this way, the complexity of the optimization task has already been reduced by one degree. Of course, this does not concern the single filter case.

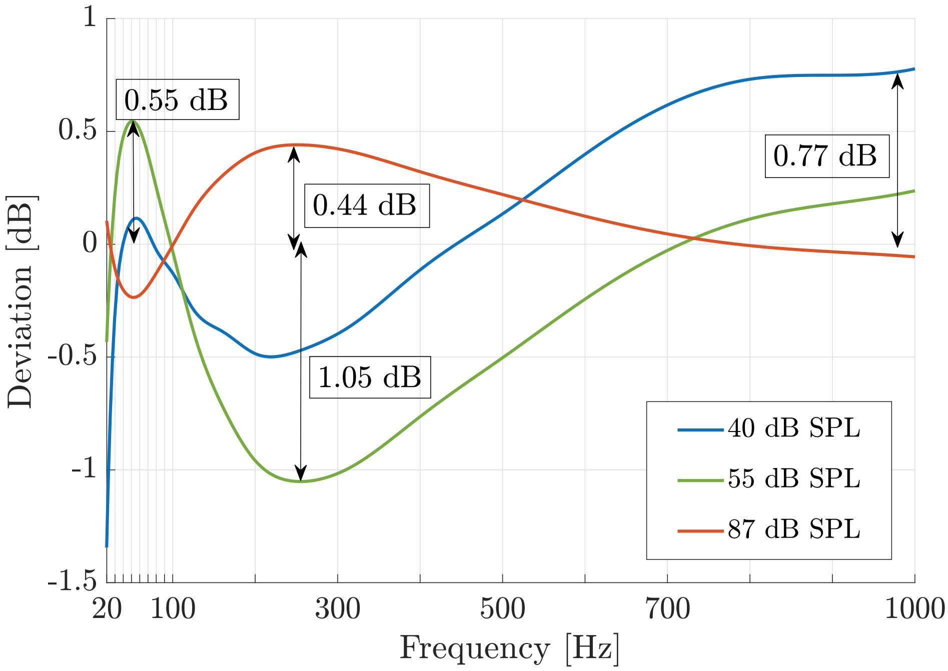
The crucial step in optimization seems then to be the choice of the poles, i.e. the crossover frequencies. It is easy to change parameters of a digital filter when working in real-time applications; however, given the (reasonable) short range of considered mastering levels and the definite frequency span of the trace-guide, fixing the poles simplifies the problem even further and without loss of generality, leaving only G to be modified as the playback level changes.

Multiple GA runs return, as consistent optimal solution for the poles:

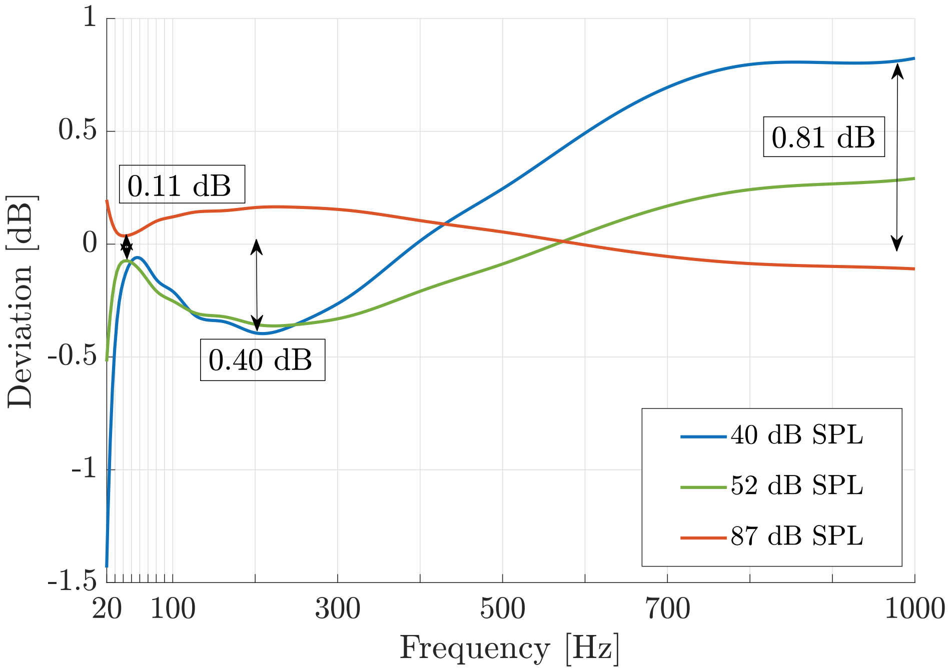
- $f_c = 122.0552$ Hz, for the first-order shelving filter;
- $f_{c1} = 61.0810$ Hz, $f_{c2} = 242.0095$ Hz, for the cascade of first-order shelving filters.

Maximum deviations from trace-guide are reported in Fig. 3.9. As it can be seen, both cases provide interesting results, with small errors for all considered listening levels and minimum error on the mid frequencies.

To achieve high-fidelity, a maximum deviation of ± 1 dB from the trace-guide is desired. As shown in Fig. 3.9b, the filter cascade error always lies inside such a range, while the single filter deviation slightly exceeds -1 dB around 250 Hz (Fig. 3.9a). Although, the cascade already satisfies requirements, the use of a single filter is desirable to further reduce complexity and computation time.



(A) Single first-order filter



(B) Cascade of first-order filters

FIGURE 3.9: Worst-case deviations from trace-guide given by first-order low-shelving filters, with peaks of max deviation.

3.4.3 Gain adjustment

Since the crossover frequency has been optimized and fixed, a possible solution is to adjust the gain with a small offset α in order to compensate for the deviation peak around 250 Hz, without exceeding the range somewhere else.

The modified filter gain is then determined as:

$$G_{\mathrm{dB}} = \Delta L_p(20, L_M, L_L) + \alpha. \tag{3.6}$$

Running the GA again, an optimal coefficient $\alpha = 0.485\,\mathrm{dB}$ is found. Fig. 3.10 shows that this small variation produces an appreciable reduction of the deviation peak, with acceptable degradation of the deviation from the trace-guide in the rest of the bass range. It is interesting to notice how the regions in the spectrum where the deviation peaks lie correspond approximately to the frequencies of the optimized poles for the cascade of first-order shelving filters.

The magnitude responses of the first-order shelving filters using the modified gain and (3.2) is shown in Fig. 3.11. It is interesting to confront it with Fig. 2.3: the resemblance of such responses to the ideal compensation curves is easily acknowledgeable, with a major difference towards the lowest frequencies (as it was said, this is actually desirable to avoid instabilities and noise amplification). It must be underlined that a 1 dB change in the applied gain does not correspond to a 1 dB SPL change in the level difference, as this might not be trivial.

The filter coefficients (with $a_0 = 1$) used for these curves are listed in Table 3.1.

All coefficients are real and next to the unit value. This translates into poles and zeros which are really close to each other and near the unit circle on the complex plane (Fig. 3.12). As it can be seen, the pole and the zero grow close to each other as the difference between mastering level and listening level decreases, up to when they coincide. Above that point, i.e. when the listening level is higher than the mastering one, the pole is greater than the zero and they start to grow apart.

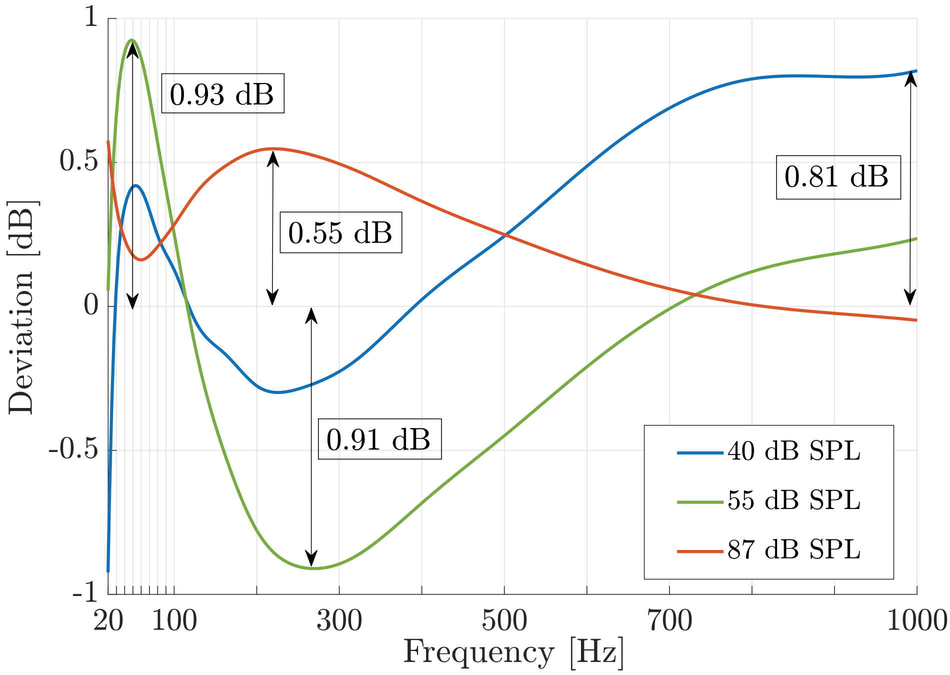


FIGURE 3.10: Deviation from trace-guide given by first-order low-shelving filter, with gain adjustment $\alpha = 0.485$ dB.

L_L	Numerator		Denominator
[dB SPL]	b_0	b_1	a_1
90	0.9952	-0.9821	-0.9773
80	1.0005	-0.9827	-0.9832
70	1.0058	-0.9818	-0.9876
60	1.0117	-0.9791	-0.9908
50	1.0186	-0.9746	-0.9932
40	1.0271	-0.9678	-0.9949

TABLE 3.1: Shelving filter coefficients for various L_L, $L_M = 80$ dB.

It is valuable to notice that, due to the gain adjustment, the response at L_M is not transparent anymore. However, the separation between the zero and the pole is so small that any difference in the sound reproduction introduced by the first-order single shelving filter in such a case will fall below just noticeable difference (JND), so the solution is still correct. Eventually, it can be possible to stop the real-time filtering process (i.e. put $b_0 = b_1 = a_0 = a_1 = 1$) if the detected listening level is found equal to L_M.

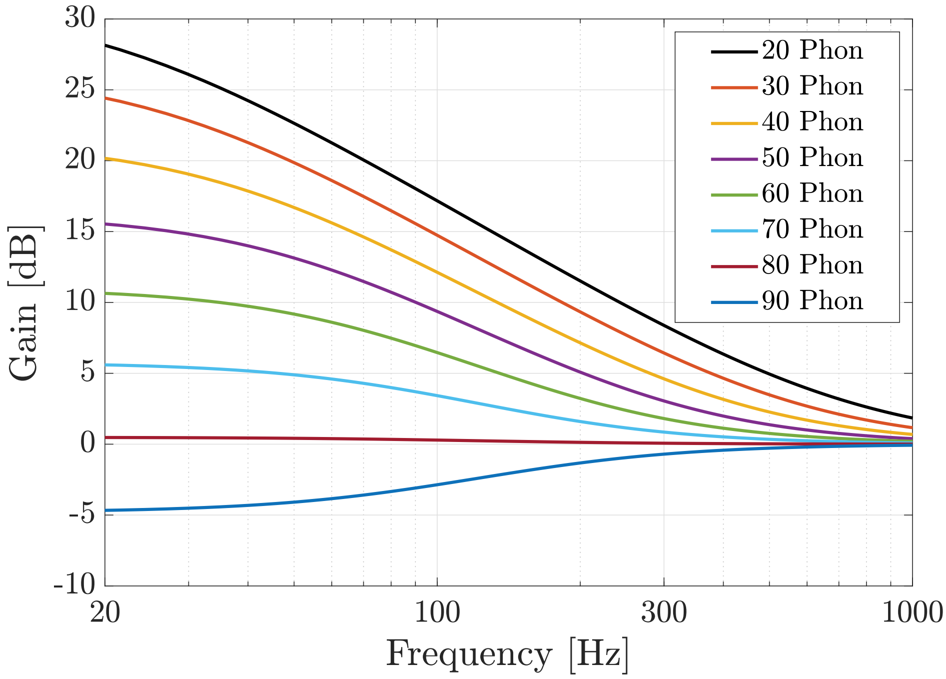


FIGURE 3.11: Magnitude responses of the first-order shelving filter at $f \leq 1$ kHz for certain listening levels, when $L_M = 80$ dB.

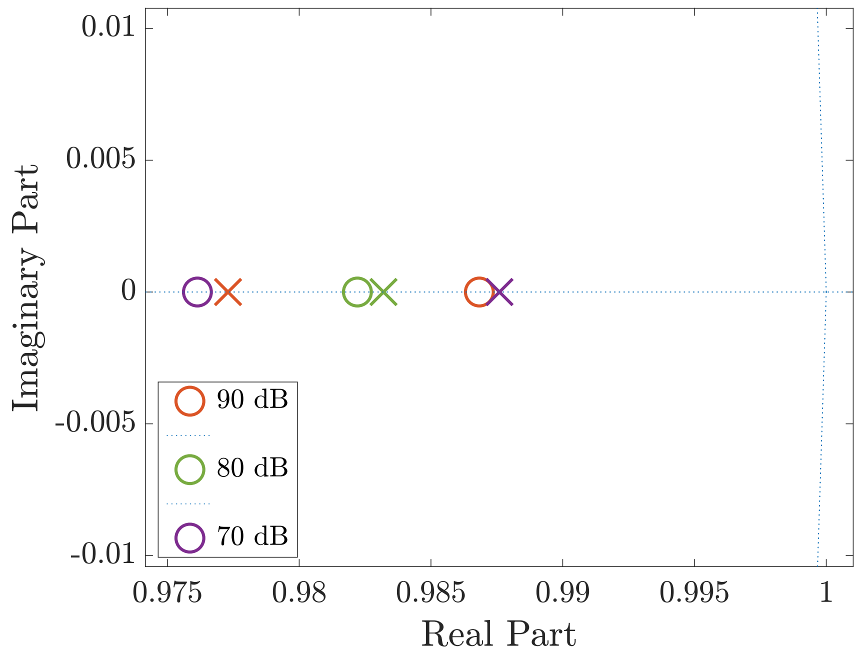


FIGURE 3.12: Pole/zero couples of the shelving filter on the complex plane.

Chapter 4

Listening Test

In the field of human perception and especially in psychoacoustics, subjective listening test are time-consuming and often difficult to evaluate, but they can provide meaningful data and useful suggestions when designed and conducted in the proper way. In this chapter, a custom listening test validating the proposed compensation technique is described. Details about method, environment, setup and results are reported in the following sections.

4.1 Design

It is clear that classical MUSHRA or ABX tests are not well suited for validating the proposed method for loudness compensation; therefore, an ad-hoc design has to be preferred. After a thorough literature review, no example of previous listening test for similar applications has been found, although the use of custom ones has been sometimes reported. A new test has then to be developed, under the following requirements:

- *blinded* test, which means that stimuli must be unknown to subjects (*single-blind*) and eventually also the researchers (*double-blind*);
- $N \geq 10$, where N is the number of participants, to have meaningful statistical analysis;
- limited subject knowledge of the test, to avoid biasing.

4.1.1 Subjects and environment

The test was conducted on a selection of experienced listeners among Aalto Acoustics Lab staff member. Prior to the test, each one of them signed a declaration, stating that they have no known hearing impairment or other medical condition and that they accept the use and share of collected data under anonymous terms.

The test was designed and conducted in MATLAB environment on a MacOS machine, using a pair of Sennheiser HD 650 (Fig. 4.1) dispatched inside a listening booth at the Aalto Acoustics Lab which provides sound isolation via 40 dB room absorption. Technical features of the headphones are reported in Tab. 4.1.

FIGURE 4.1: Picture of Sennheiser HD 650.

Headphone type	Circumaural
Frequency response	10 Hz–41 kHz
Total harmonic distortion (THD)	< 0.05 %
Contact pressure	$\sim$2.5 N
Jack plug	6.35mm, stereo, gold
Cable length	3m
Weight	260g
Nominal impedance	300 Ω

TABLE 4.1: Technical features of Sennheiser HD 650.

4.1.2 Sounds

Audio samples were chosen from different genres for having a prominent bass line and other different spectral features. They consisted of short tracks (4 to 8 seconds) cut from the following songs:

1. Queen, "*Another One Bites The Dust*" (1980);

2. White Stripes, "*Seven Nation Army*" (2003);

3. Daft Punk, "*Around The World*" (1997).

From further on, each track will be identified with its number from the list above, e.g. Track 1, Track 2, Track 3. Track 2 is composed by just bassline and drumline, showing narrow spectral content concentrated in the bass range. Track 3 present a broader spectral content; same for Track 1, which also includes vocals.

Three samples are enough for the task, since each one of them has repetitions during the test. The singular spectral behaviour they present allow to compare the effects of the applied loudness compensation for different situations during music listening and have a wider and more heterogeneous set of results.

Tracks have been cut in such a way that an infinite loop is possible without artifacts or harsh discontinuities in sound reproduction. Audio editing has been done on an Ubuntu 18.04 machine using *Audacity*® (Fig. 4.2).

4.1.3 Test modality

Subjects under test are presented with seven instances of each track, for a total of 21 audio samples and 21 test steps. Each step holds a version of the track played at L_M, named *reference*, and an attenuated variant, representing the playback.

Applied reduction of loudness varies between 0 and 40 dB in steps of 10 dB, corresponding to five different listening levels in the 40–80 dB SPL range. 80 dB SPL

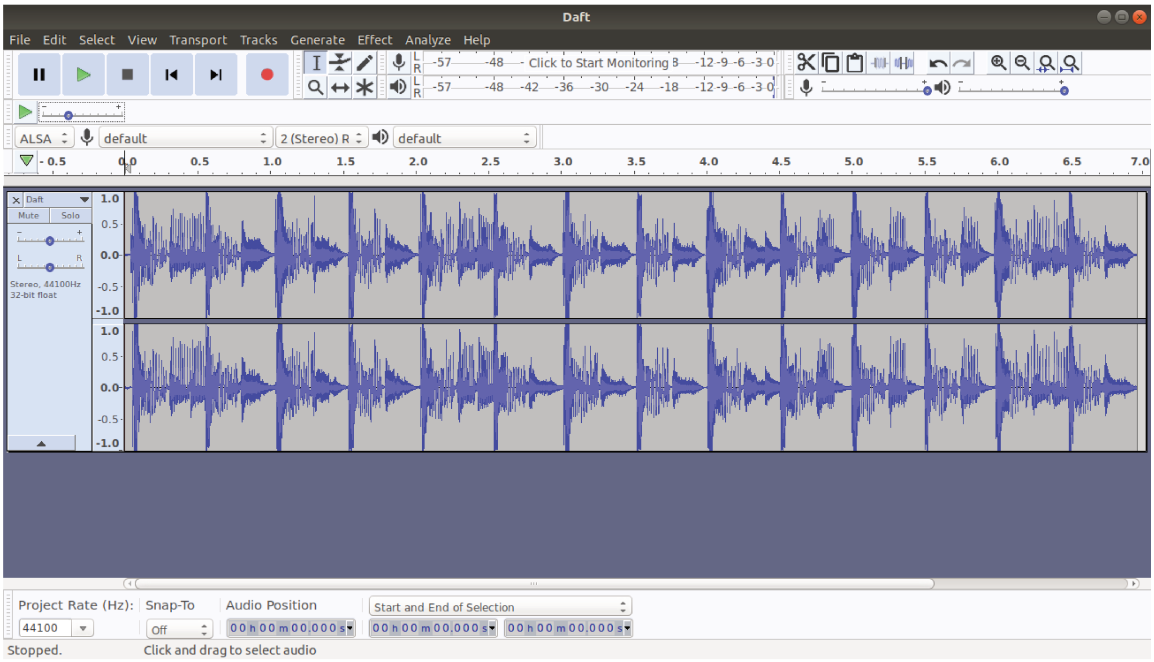


FIGURE 4.2: Audio editing in *Audacity*®.

(no reduction, same as reference) and 60 dB SPL (20 dB attenuation) are presented twice per each track: such repeated plays are used during the screening phase to evaluate subject consistency and then discarded before the actual statistical analysis of results.

Loudness compensation is applied to the attenuated variant using a single first-order low-shelving filter, whose crossover frequency is fixed at 122.0552 Hz (see Sec. 3.4.2) and whose gain is modified by subjects during the test, as they try to adapt to different listening levels.

A critical aspect for the sake of the test is the *a priori* knowledge of participants. Subjects are asked to focus on the bass of the played songs, in particular on the balance between the overall loudness and the bass loudness of the reference, and to apply a compensation in order for the spectral balance of the variant to match the reference balance, but at a different level.

Subjects have access to a horizontal slider (Fig. 4.3) that –it is told them– allows to give “*boost or reduction*” to the bass of the variant audio sample. The range of the

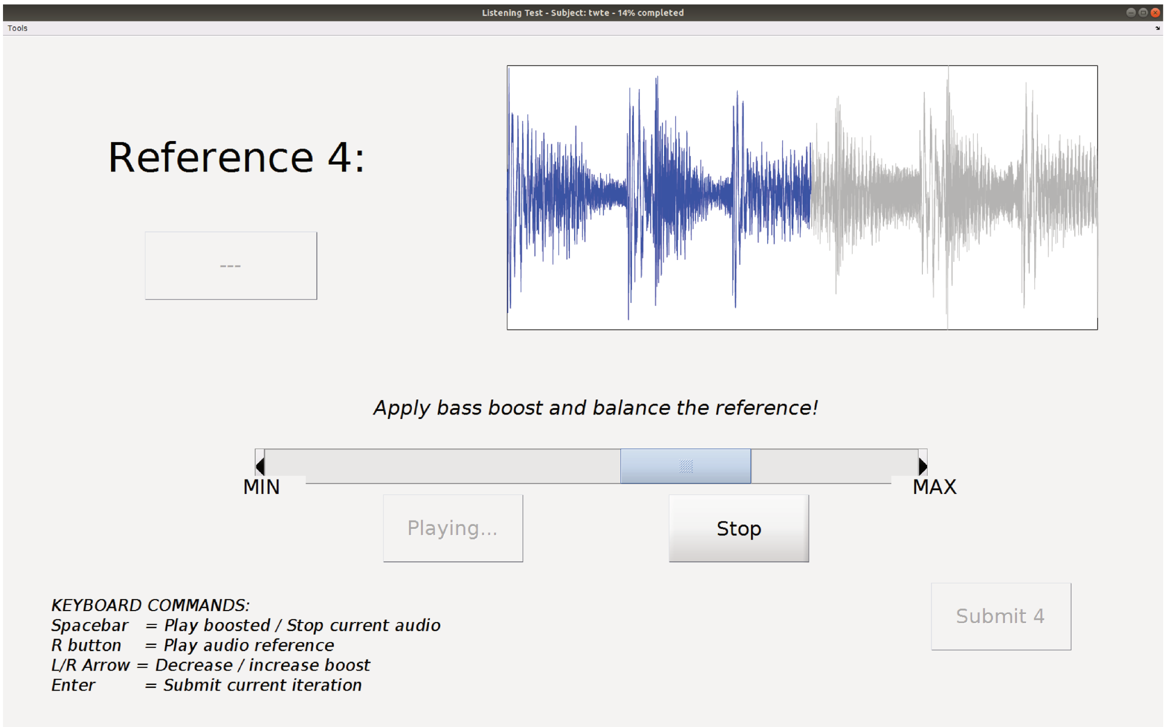


FIGURE 4.3: Screen-shot of the test GUI.

slider is hidden and slightly randomized at each test step, showing only the labels "*Min*" and "*Max*" at its two extremes.

The slider actually selects a different gain for the filter from the ELLC trace-guide (Sec. 3.4). Slider movements are discretized and each step corresponds to a 2-Phon variation in the selection of the curve. If the randomization brings the slider interval out of the loudness range described by the ELLC, saturation to highest or lowest value is introduced.

Audio samples are set to play in a continuous loop until they are manually stopped. While it is possible to reproduce the variant as many times and as long as desired, the maximum play count of the reference is limited to two, the first starting automatically at each new step of the test. This means that, after stopping the reference the first time, subjects are allowed to play it again just one more time. This choice has been made to avoid the listener to go "back and forth" from the reference to the variant and force them to pay extra focus on the task.

Given the fast decay of human memory of sounds, the subjects are suggested to get a general idea of the frequency components of the reference during the first play, then to reproduce the variant and explore the amount of possible “boost” given by the slider, before getting back to the reference and gain a more clear sense of the spectral balance. After the second stop, a final choice for variant compensation should be made.

Listeners are given a 3-step training session before starting the actual test to get acquainted with the interface, the keyboard shortcuts and the task itself. The results of the training session are not considered in the statistical analysis.

4.1.4 Evaluation method

The way in which the statistical analysis is going to be conducted and how data are going to be evaluate is critical. It could happen that bad results are misinterpreted as good, or, conversely, good results are mistaken for bad, due to poor methods of evaluation.

The idea behind the listening test is to verify the ability of the proposed filter to match the imbalance and consequently the needed compensation perceived by a common listener. Difference in subjective hearing and the difficulty of the task should be taken into account. For this reasons, results are considered “good” if the median of the collected data, for each audio sample at each level, is close enough to the real value required by the ideal compensation function described in this thesis, while keeping the variance as low as possible without general restrictions. Louder levels are expected to produce low variance (ideally zero variance at reference level), while quieter levels are inevitably going to give worst results.

A good method for evaluation of collected data is *boxplot*. A boxplot graphically depicts groups of numerical results through their quartiles (Fig. 4.4). It produces a box, containing 50% of the data, and a line inside it, representing the median. There are lines extending vertically from the boxes, called *whiskers*, indicating variability outside the upper and lower quartiles. The end of the whiskers represent the lowest datum still within 1.5 IQR (*Inter-Quartile Range*, equal to the difference between 75th and 25th percentiles) of the lower quartile, and the highest datum still within 1.5 IQR

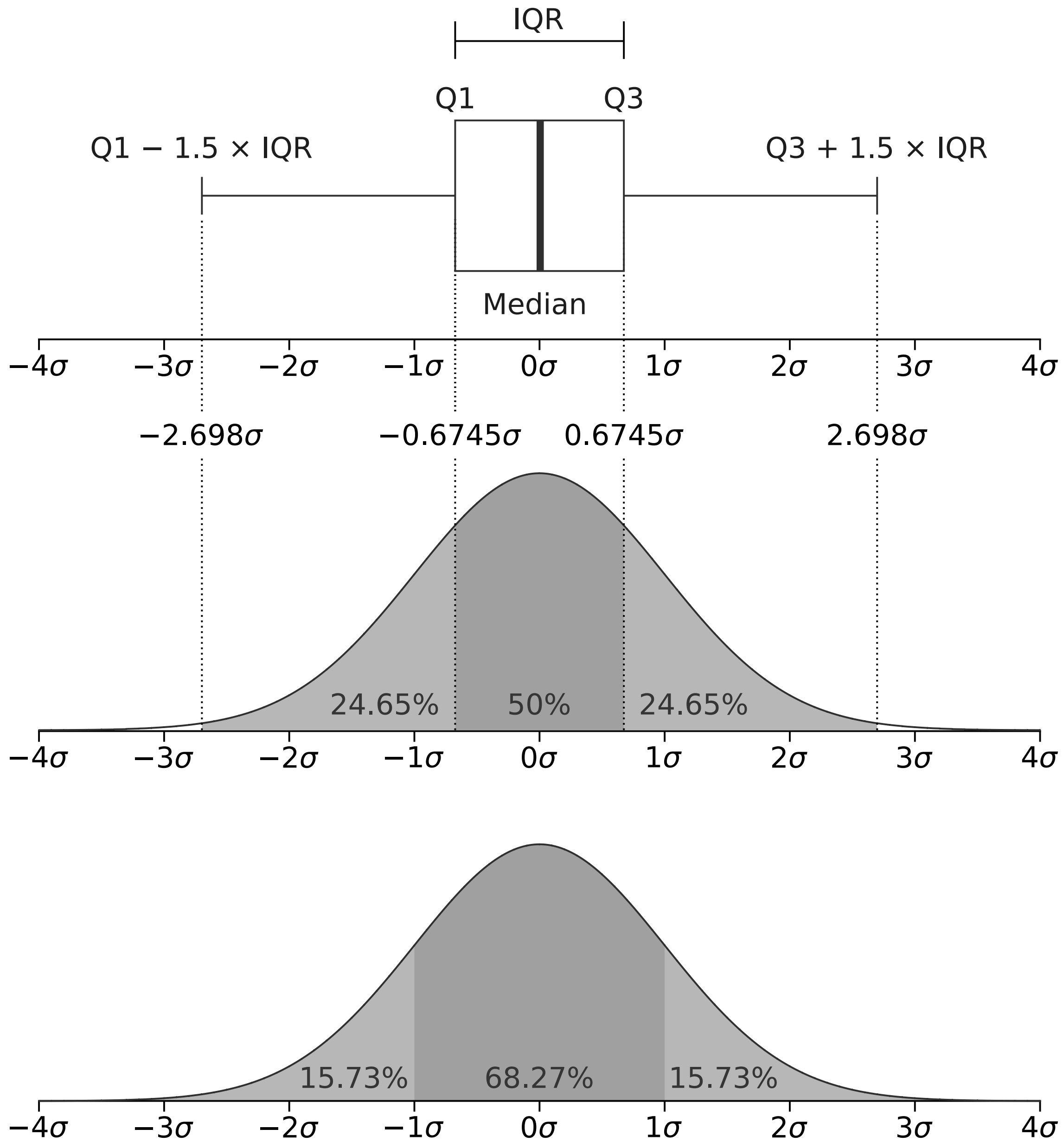


FIGURE 4.4: Example of a boxplot and comparison with the probability density function.

of the upper quartile. *Outliers* (anomalous values) are plotted as individual points.

The spacings between the different parts of the box indicate the degree of dispersion and asymmetry in the data.

Boxplots are non-parametric: they display variation in samples without making any assumptions of the underlying statistical distribution, which comes in handy in situations with a low-number statistical population.

Boxplots allow a quick graphical examination of the dataset, and have some advantages with respect to a classical histogram or kernel density estimate, in particular for comparing distributions between several groups of data.

4.2 Level calibration

An accurate measurement of the loudness level is critical for the quality and validity of the test, so a calibration phase has to be performed. Used instrumentation involves a RME Fireface 800 and a G.R.A.S. 45CA Headphone Test Fixture in compliance with the IEC 60318-4:2010 occluded-ear simulation [10] in order to measure the audio tracks involved in the test and get a precise estimation of their loudness.

The different samples, played through the headphones allocated on the ear simulator, are first loudness matched by using ITU-R BS.1770-4 [14] loudness measure, with a tolerance of ± 0.5 LKFS. Their playback level is then set to be approximately 80 dBA by acting on sound card gain.

The actual measured dBA can differ, since the levels depend on the contents of the considered track. Measured ITU-based loudness levels and maximum A-weighted dB levels are reported in Table 4.2.

The calibration phase is conducted via MATLAB environment on a MacOS machine. The response of the ear simulator has been recorded and used inside a custom script to normalize acquisitions of loudness and A-weighted sound pressure levels.

A picture of the calibration setup is shown in Fig. 4.5.

	Loudness [LKFS]	SPL_{MAX} [dBA]
Track 1	-11.2752	83.3671
Track 2	-11.0675	76.9724
Track 3	-10.9571	81.1604

TABLE 4.2: ITU-based loudness in loudness units relative to full scale (LKFS) according to ITU-R BS.1770-4 and the maximum SPLs (in dBA).

FIGURE 4.5: Calibration setup: MacOS machine, RME Fireface 800 and G.R.A.S. 45CA disposed on a moving table.

4.3 Pilot test

A pilot experiment is quite important to highlight:

- any practical problems;
- the suitability of the used audio sample;
- the amount of time required to perform the test.

The third point is extremely critical. A long test usually tires the listener, making them incapable of fully focusing on the task and uncaring of the quality of their answers, just to end the test session as soon as possible. Vice versa, a short test is not able to provide meaningful results for the statistical analysis.

Two different pilot tests have been conducted prior to the final session session with the targeted subjects. The initial version of the listening test featured 40 audio samples and a fixed interval of boost in the slider, with a 3-Phon change for each step. It proved to be too long and level patterns are easily identifiable on the slider, so in the final version those features were changed as described above.

4.4 Screening

In order to discard inconsistent listeners, 80 dB SPL and 60 dB SPL cases are presented twice. For screening, it is not important how accurate subjects are, but their degree of repeatability. For this reason, the absolute difference between the first and the repeated value was calculated at both levels, for each song and for every subject. A double-threshold method has been designed to evaluate consistency:

1. if $\forall i \; \Delta 80_i \leq 6\,\text{dB}$, subject is consistent;
2. otherwise, if $\Delta 80_i > 6\,\text{dB}$ for one track and $\Delta 60_i \leq 10\,\text{dB}$ for at least two tracks, subject is consistent;
3. otherwise, subject is inconsistent and discarded.

Here, $i = 1, 2, 3$ is the number of the track and $\Delta 80_i$ and $\Delta 60_i$ are the differences of the two instances. Since human perception is evaluated, it is reasonable to have a stricter threshold at the reference level (80 dB), where the spectral balance of two signals with the same level are matched, and a more relaxed threshold for the attenuated level (60 dB), when task of matching the spectral balance of two signals with different listening levels is harder. A total of 18 subjects participated in the test; 11 of them passed the consistency screening and are included in the statistical analysis.

4.5 Results and analysis

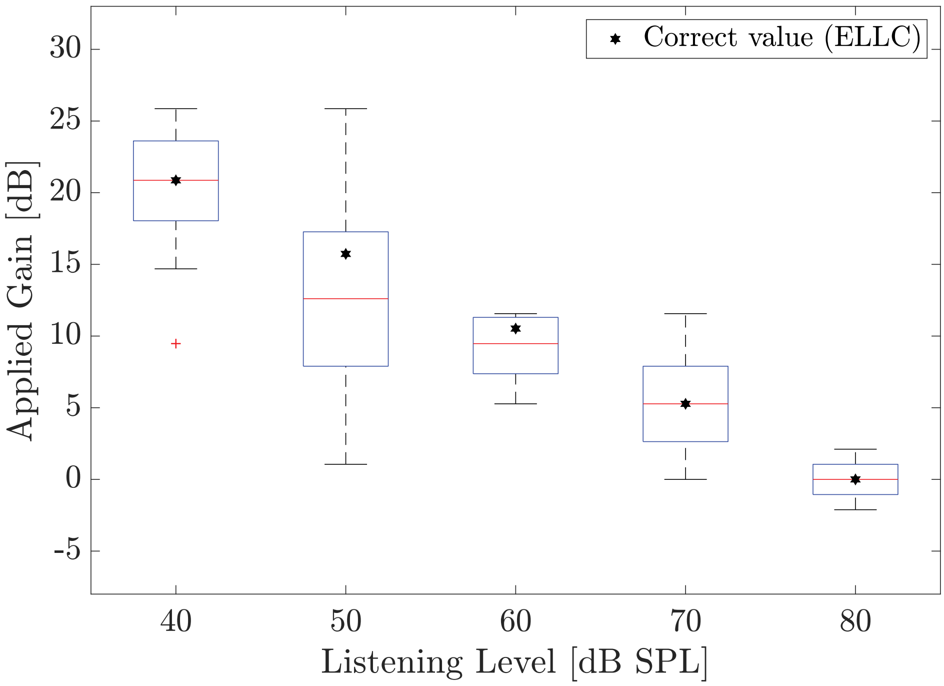
Results from the listening test show that participants were able to follow the compensation trend set by the equal-loudness contours. This is shown in Fig. 4.6 and 4.7, where the boxplots present the median (red line), the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles (blue box), the extension to the most extreme data points not considered outliers (black whiskers) and the outliers (red cross). Furthermore, the black markers represent the correct level compensation required according to the equal-loudness contours.

Applied compensation versus the listening level of the track is plot in Fig. 4.6, showing how such compensation adapts to the level and increases towards the lowest levels. Level deviations are reported in Fig. 4.7, where the goodness of the results and the cases of under or over compensation are easier to be seen.

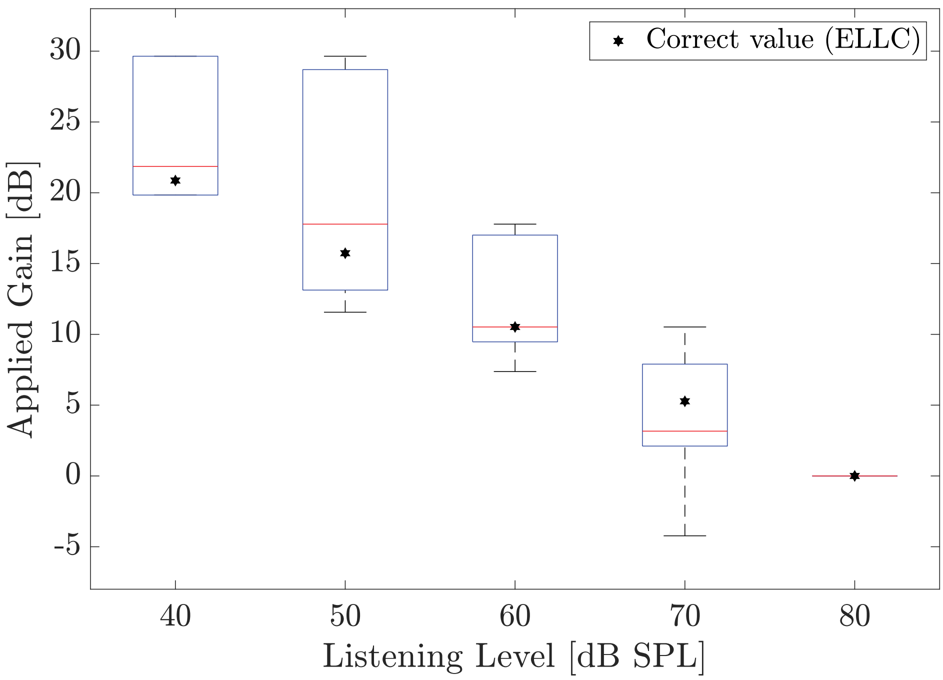
As expected, data presents moderate variance, due to the difficulty of the task; nevertheless, the median of the error in level evaluation always lies in a close range near 0 dB (Fig. 4.7), satisfying what was required beforehand.

Taking a further analysis of the results, it is possible to state that:

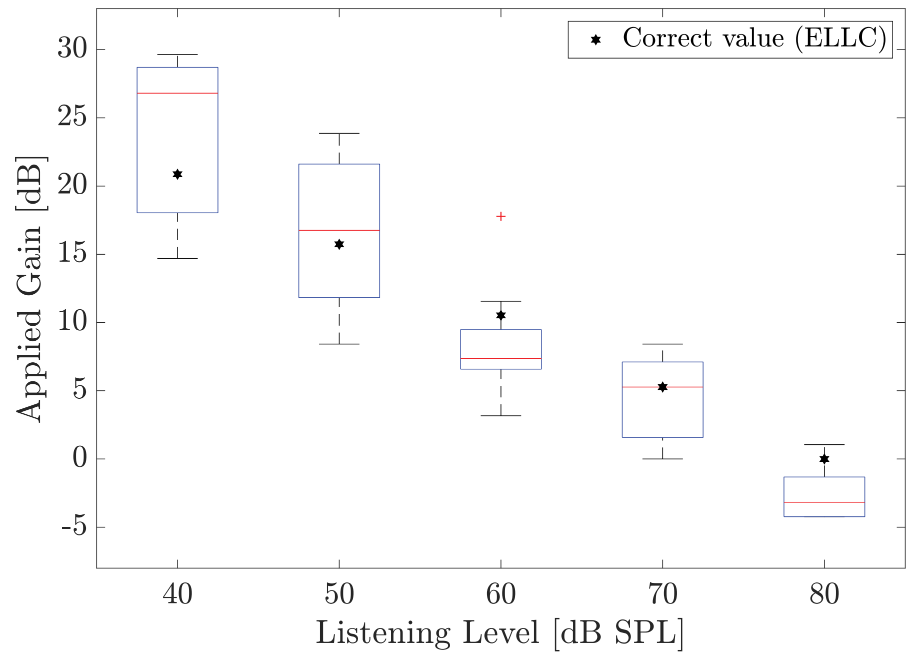
- the reference was matched quite well by almost all listeners for all samples, with slightly worst accuracy for Track 3 (Fig. 4.6c and 4.7c);
- fairly good results were obtained in typical music listening range (60 and 70 dB SPL), which was the focus of this method;
- the variance increased towards the lowest levels (40 and 50 dB SPL), where sound was really quiet and the task of matching the perceived spectral balance became harder.



(A) Track 1

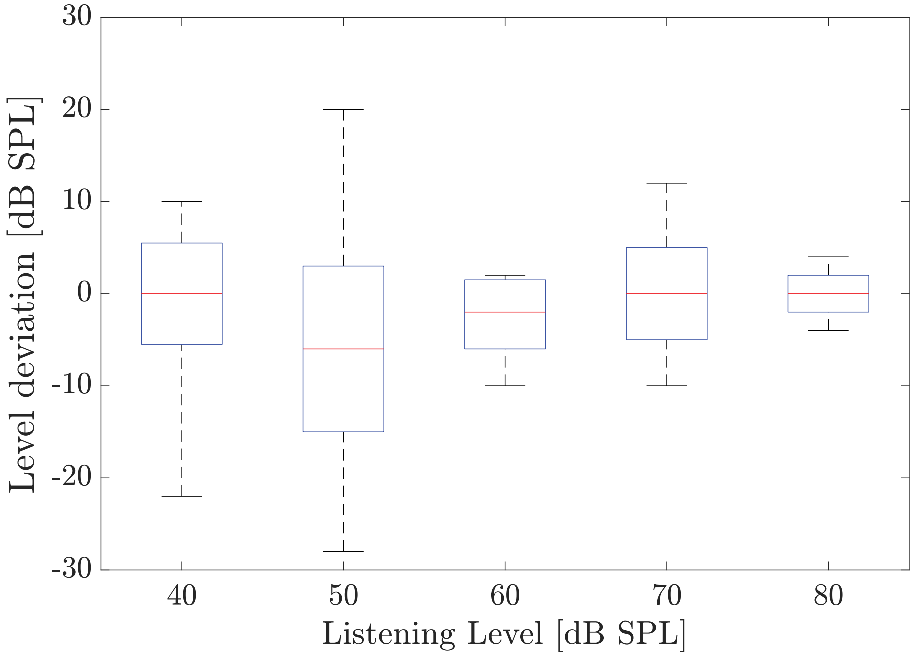


(B) Track 2

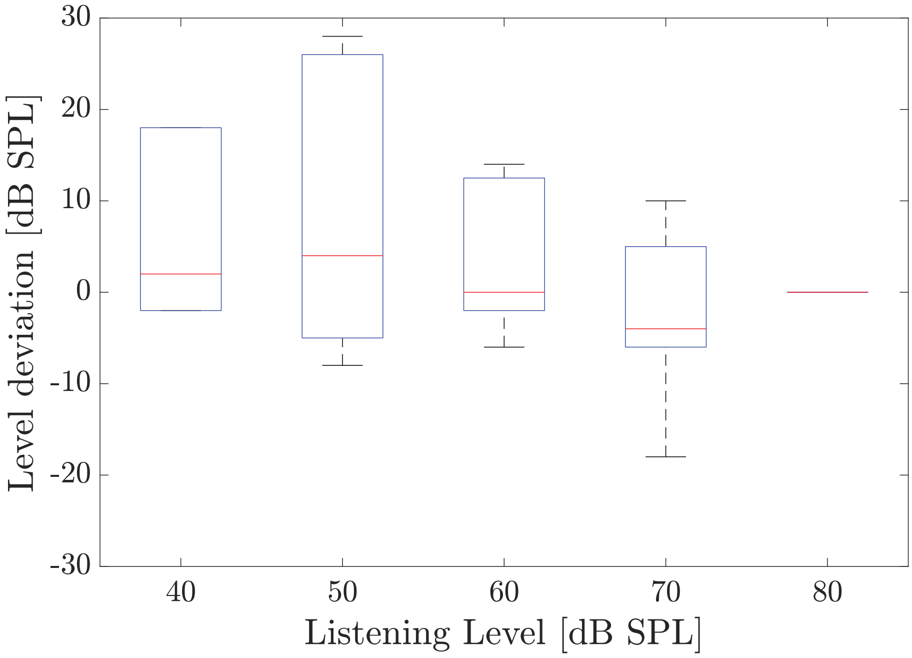


(C) Track 3

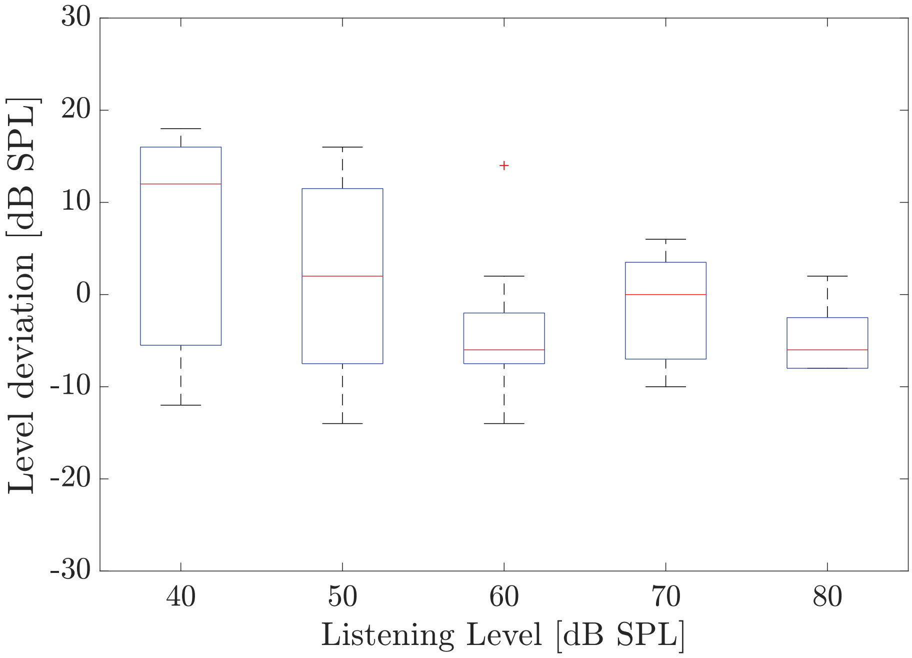
FIGURE 4.6: Test results, grouped by song. Boxplot of compensation introduced by subjects at each listening level.



(A) Track 1



(B) Track 2



(C) Track 3

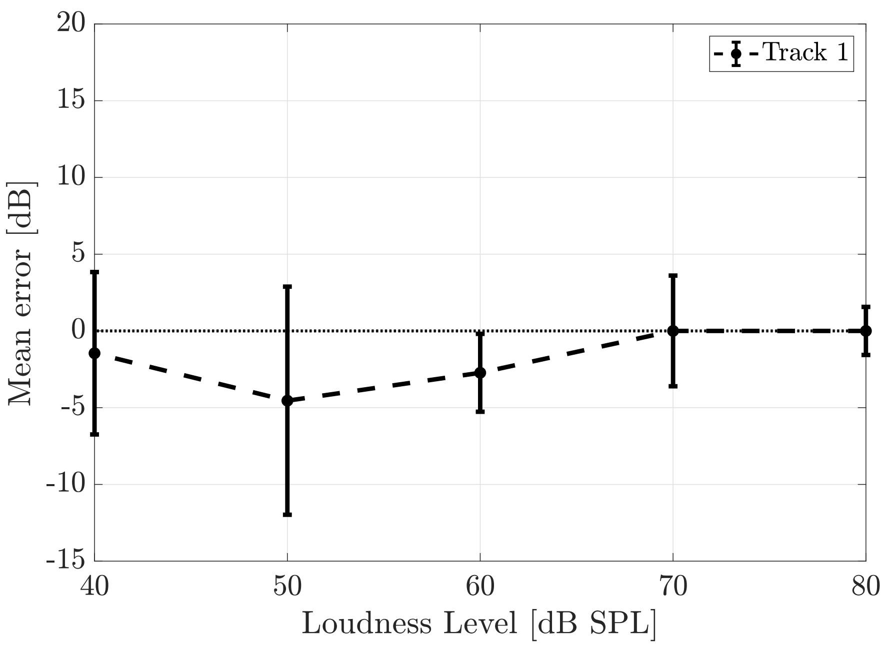
FIGURE 4.7: Test results, grouped by song. Boxplot of error in level evaluation versus the corresponding correct level.

It is interesting to notice that, for Track 2 (Figs. 4.6b and 4.7b), the majority of the listeners tended to over-compensate when the music level went down. This makes sense, due to the sample having narrow spectral content and, as a consequence, no "untouched" frequency components to be compared to, increasing the difficulty.

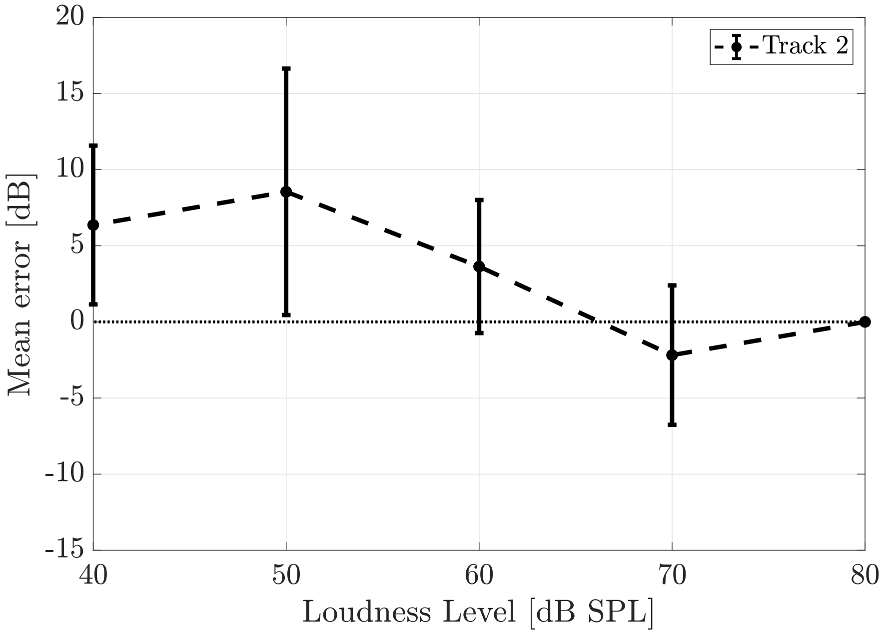
After taking the test, the subjects were asked for a feed-back. They confirmed the difficulty of matching the spectral balance, when the level of reproduction went down. It was also difficult to notice the audible difference among small changes of the slider, since only the lowest frequencies were affected. They also stated that bass contribution was noticeable and pleasing.

As a overall, it can be said that the conducted listening test has validated the proposed compensation method.

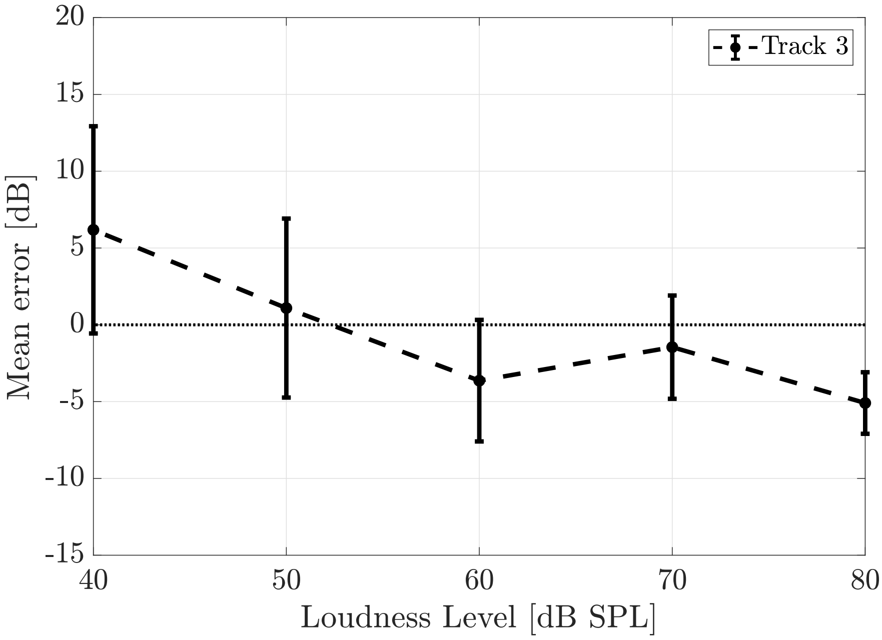
To complete the statistical analysis of the results, Fig. 4.8 shows the mean values of the errors, with 95% confidence intervals.



(A) Track 1



(B) Track 2



(C) Track 3

FIGURE 4.8: Test results, grouped by song. Mean error values of the error for each sample, with 95% confidence intervals, versus the corresponding correct level.

Summary and conclusions

A loudness compensation function derived from the equal-loudness-level contours and implemented via digital filters was proposed. Such function introduces an adaptive contribution to the bass based on the listening level, in order to balance the perceived spectral variations given by the nonlinear response of the human hearing system.

Among different typologies, first order low-shelving filter with gain adjustment and fixed crossover frequency was shown to provide a high-fidelity approximation of the compensation function for wide range of listening levels. Low complexity and widespread usage enable a real-time implementation.

Results from a listening test conducted on a set of trained listeners showed the goodness of the compensation method, in particular for what concerns the typical range of listening levels for music.

Future work might include on-chip applications, customization for specific hardware or environments and new listening tests, conducted on larger pools of non-trained listeners reflecting consumer market. It might be also interesting to evaluate the effect of masking on the application of this method and its potential use.

Appendix A

Code implementation of the ELLC

This appendix contains the implementation of the equal-loudness level contours and the proposed adaptive compensation function in MATLAB.

A.1 Equal-loudness level contours

The standard curves as described in ISO 226:2003 [12].

```
function [spl, f] = ellc(phon)

% Written by Leonardo Fierro, Oct. 4, 2018

% Data setting
% f = frequency; Lp = Sound Pressure Level; Ln = Loudness level;
% af = exponential for loudness perception;
% Tf = threshold of hearing;
% Lu = magnitude of linear transfer function, norm. at 1 kHz

f = [20 25 31.5 40 50 63 80 100 125 160 200 250 315 400 500 ...
    630 800 1000 1250 1600 2000 2500 3150 4000 5000 6300 8000 ...
    10000 12500];
```

```
af = [0.532 0.506 0.480 0.455 0.432 0.409 0.387 0.367 0.349 ...
 0.330 0.315 0.301 0.288 0.276 0.267 0.259 0.253 0.250 0.246 ...
 0.244 0.243 0.243 0.243 0.242 0.242 0.245 0.254 0.271 0.301];

Lu = [-31.6 -27.2 -23.0 -19.1 -15.9 -13.0 -10.3 -8.1 -6.2 -4.5 ...
    -3.1 -2.0 -1.1 -0.4 0.0 0.3 0.5 0.0 -2.7 -4.1 -1.0 1.7 ...
     2.5 1.2 -2.1 -7.1 -11.2 -10.7 -3.1];

Tf = [78.5 68.7 59.5 51.1 44.0 37.5 31.5 26.5 22.1 17.9 14.4 ...
     11.4 8.6 6.2 4.4 3.0 2.2 2.4 3.5 1.7 -1.3 -4.2 -6.0 -5.4 ...
    -1.5 6.0 12.6 13.9 12.3];

if((phon < 0) | (phon > 90))
    disp('Cannot-evaluate-SPL-level-(Out-of-bounds)')
    spl = 0;
    f = 0;
else
    Ln = phon;

    % Formula from ISO 226 Section 4.1
    Af = 4.47e-3 * (10.^(0.025 * Ln) - 1.15) + ...
        (0.4 * 10.^(((Tf + Lu)/10) - 9)).^af;
    Lp = ((10./af) .* log10(Af)) - Lu + 94;

    spl = Lp;
end
```

A.2 Interpolation and compensation function

Equal-loudness contours interpolation, difference curve and compensation function, as proposed and described in Chapter 2.

```
function [filtTrace,f] = fitLoudnessContoursX(dB);

% Written by Leonardo Fierro, Oct. 5, 2018

%% Calculate SPLs
phon = 20:10:90;
[spl,f] = ellc(phon');

%% Linear interpolation
phonFit = 20:1:90;
Npoints = size(spl,2);
spl_fit = zeros(length(phonFit),Npoints);
for i = 1:Npoints
    v = spl(:,i)';
    vq = interp1(phon,v,phonFit);
    spl_fit(:,i) = vq';
end

%% Difference from mastering
masterCurve = spl(end,:);
lLevel = find(phonFit == dB);
listenCurve = spl_fit(lLevel,:);

diffCurve = masterCurve - listenCurve - 80 + dBChoice;
filtTrace = -diffCurve;
```

Appendix B

Code implementation of filters

This appendix contains the implementation of fractional order filters and low-shelving filters in MATLAB.

B.1 Fractional order filters

Implementation of a first "order" fractional order filter.

```
function [num,den] = fof1(Go,pol,Fs)

% INPUTS
% Go = Gain at low frequencies (dB)
% pol = pole frequency
% Fs = sample frequency
% OUTPUTS
% num = numerator coefficients b0 and b1
% den = denominator coefficients a0 and a1

% Written by Leonardo Fierro, Oct. 9, 2018.

% Zero derivation
G = 10^(Go/20);
```

```
zer = round(G*pol);

% Transfer function coefficients
a0 = (1+pol(1)/Fs);
a1 = (1-pol(1)/Fs);
b0 = -(1+zer(1)/Fs);
b1 = (1-zer(1)/Fs);

% Transfer function polynomials
den = [a0 a1];
num = [b0 b1];
```

B.2 Shelving filters

B.2.1 First-order low-shelving filter

First order low-frequency shelving filter derived in [37].

```
function [num,den] = shelf1low(G,wc)

% INPUTS
% G = Gain at low frequencies (linear)
% wc = crossover frequency
% OUTPUTS
% num = numerator coefficients b0 and b1
% den = denominator coefficients a0 and a1

% Written by Vesa Valimaki, Nov. 5, 2015.
```

```
% Transfer function coefficients
a0 = tan(wc/2) + sqrt(G);
a1 = (tan(wc/2) − sqrt(G));
b0 = (G*tan(wc/2) + sqrt(G));
b1 = (G*tan(wc/2) − sqrt(G));

% Transfer function polynomials
den = [a0 a1];
num = [b0 b1];
```

B.2.2 Second-order low-shelving filter

Second order low-frequency shelving filter derived in [37].

```
function [num,den] = shelf2low(G,wc)

% INPUTS
% G = Gain at low frequencies (linear, not dB)
% wc = crossover frequency (radians)
% OUTPUT
% num = numerator coefficients b0, b1, and b2
% den = denominator coefficients a0=1, a1, and a2

% Written by Vesa Valimaki, 29 April 2016

% Filter coefficients
Omega = tan(wc/2);
a0 = sqrt(1/G)*Omega^2 + sqrt(2)*Omega*G^(−1/4) + 1;
a1 = 2*(sqrt(1/G)*Omega^2 − 1);
a2 = sqrt(1/G)*Omega^2 − sqrt(2)*Omega*G^(−1/4) + 1;
```

```
b0 = sqrt(G)*Omega^2 + sqrt(2)*Omega*G^(1/4) + 1;
b1 = 2*(sqrt(G)*Omega^2 - 1);
b2 = sqrt(G)*Omega^2 - sqrt(2)*Omega*G^(1/4) + 1;

% Transfer function
den = [a0 a1 a2];
num = [b0 b1 b2];
```

Appendix C

Code implementation of the listening test

This appendix contains the implementation of the most significant Matlab functions used to develop and run the described listening test.

C.1 Real-time filtering during playback

Callback associated to "Play" button on test GUI, creates an infinite loop until "Stop" button on test GUI is pressed. PlayRec libraries and proposed filtering are used.

```
function playFiltered_button_Callback(hObject, eventdata, handles)

% Written by Leonardo Fierro, Dec. 27, 2018

% Return focus to GUI
set(hObject, 'Enable', 'off');
drawnow;
set(hObject, 'Enable', 'on');

handles.status.isplaying = true;
set(handles.playReference_button,'String','---');
```

```
set(handles.playFiltered_button,'String','Playing...');
handles = update_GUI(handles);
guidata(hObject, handles);
z = zeros(1,2); % delay handler for filter function

while(handles.status.isplaying) % loop for paging and buffering of input and output
    pNum = playrec('play',handles.sound.Out,handles.sound.chanListPlay);

    % Plot update
    axis(handles.axes2);
    children = get(handles.axes2, 'children');
    if length(children) > 1
        delete(children(1));
    end

    hold on

    if ((handles.sound.pageCount * handles.sound.pageSize) <= length(handles.
        ↪ sound.AudioFile)) && (handles.sound.pageCount ~= 0)
        plot(handles.axes2,handles.sound.t(1:handles.sound.pageCount * handles
            ↪ .sound.pageSize), ...
            handles.sound.AudioFile(1:handles.sound.pageCount * handles.
                ↪ sound.pageSize), ...
            'color','b');
    else
        handles.sound.pageCount = 0;
    end
```

```
handles.sound.pageNumList = [handles.sound.pageNumList pNum]; % new
    ↪ page number list
if(handles.sound.firstTimeThrough) % First time, reset the skipped sample count
    playrec('resetSkippedSampleCount');
    handles.sound.firstTimeThrough = false;
else
    if(playrec('getSkippedSampleCount'))
        %fprintf(' %d samples have been skipped, upon replay.\n',playrec('
            ↪ getSkippedSampleCount'));
        %return %Let the code recover and then reset the count
        handles.sound.firstTimeThrough = true; % start again if skipped samples
    end
end

if(handles.sound.runMaxSpeed) % max speed case (continuous poll)
    while(playrec('isFinished',handles.sound.pageNumList(1))==0)
        ;
    end
else % not max speed case (wait page finished)
    playrec('block', handles.sound.pageNumList(1));
end

handles.sound.In = handles.sound.fileReader();
guidata(hObject, handles);

if ~isempty(handles.sound.In) % if data received
    value = int8(get(handles.dBchoice_slider,'Value'));
    if mod(value,2) == 1
        value = value-1;
```

```
        end
        if value > 58
            value = 58;
        elseif value < −8
            value = −8;
        end
        index = find(handles.sound.rangeLL == (handles.sound.ML − value));

        if isempty(index) % avoid holes while sliding
            index = 1;
        end

        num1 = handles.sound.Num(index,:);
        den1 = handles.sound.Den(index,:);

        G = handles.sound.Gfix ∗ handles.sound.G;
        [handles.sound.Out,z] = filter(num1,den1,G.∗handles.sound.In,z);
    end

    if get(hObject,'enable') == string('on')
        handles.status.isplaying = false;
        break
    end

    playrec('delPage', handles.sound.pageNumList(1)); % delete page that was
        ↪ obtained
    handles.sound.pageNumList = handles.sound.pageNumList(2:end); % pop
        ↪ away page number from FIFO
    handles.sound.pageCount = handles.sound.pageCount + 1;
```

```
    drawnow;
    guidata(hObject, handles);
end

set(handles.playReference_button,'String','Play');
set(handles.playFiltered_button,'String','Play');
playrec('delPage'); % delete all pages, now loop has finished
handles = update_GUI(handles);
guidata(hObject, handles);
```

C.2 Acquisition on slider movement

Callback associated to movement on the slider on test GUI, with quantization of the step and safe margins.

```
function dBchoice_slider_Callback(hObject, eventdata, handles)

% Written by Leonardo Fierro, Dec. 27, 2018

% Return focus to GUI
set(hObject, 'Enable', 'off');
drawnow;
set(hObject, 'Enable', 'on');

% Force value to fixed range
value = int8(get(hObject,'Value'));
if mod(value,2) == 1
    value = value + 1;
    set(hObject,'Value',value);
end
```

```
% Force boundaries
Max = get(hObject,'Max');
Min = get(hObject,'Min');

if value > Max % just to be extra sure
    set(hObject,'Value',Max);
elseif value < Min
    set(hObject, 'Value',Min);
end

% Update handles structure
drawnow
handles = update_GUI(handles);
guidata(hObject, handles);
```

Bibliography

[1] Sabbir U. Ahmad. *Design of digital filters using genetic algorithms*. PhD thesis, University of Victoria, Victoria, BC, Canada, Dec. 2008.

[2] American Standards Association. *American national psychoacoustical terminology*. American national standard, March 1973.

[3] Glen Ballou. *Handbook for sound engineers*. Focal Press, 2015.

[4] Søren Bech and Nick Zacharov. *Perceptual audio evaluation-Theory, method and application*. John Wiley & Sons, 2007.

[5] Harvey Fletcher and Wilden A. Munson. Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 12(4):377–430, 1933.

[6] C. Daniel Geisler. *From sound to synapse: physiology of the mammalian ear*. Oxford University Press, USA, 1998.

[7] Oliver Hawker and Yonghao Wang. A method of equal loudness compensation for uncalibrated listening systems. In *Audio Engineering Society Convention 139*. Audio Engineering Society, 2015.

[8] Tomlinson Holman and Frank Kampmann. Loudness compensation: Use and abuse. *Journal of the Audio Engineering Society*, 26(7/8):526–536, 1978.

[9] David M. Howard and Jaime A. S. Angus. *Acoustics and Psychoacoustics*. Routledge, 5^{th} edition, 2017.

[10] IEC 60318-4. Electroacoustics - Simulators of human head and ear - Part 4: Occluded-ear simulator for the measurement of earphones coupled to the ear

by means of ear inserts. International standard, IEC/TC 29 Electroacoustics, January 2010.

[11] IEC 61672-1. Electroacoustics - Sound Level Meters - Part 1: Specifications. International standard, IEC/TC 29 Electroacoustics, September 2013.

[12] ISO 226:2003. Acoustics – Normal equal-loudness-level contours. International standard, ISO/TC 43 Acoustics, August 2003.

[13] ISO 532:1975. Acoustics – Method for calculating loudness level. International standard, ISO/TC 43 Acoustics, July 1975.

[14] ITU-R BS.1770-4. Algorithms to measure audio programme loudness and true-peak audio level. Standard, ITU/Radiocommunications, October 2015.

[15] ITU-R BS.468-4. Measurement of audio-frequency noise voltage level in sound broadcasting. International standard, ITU/Radiocommunications, 1986.

[16] Robert A. Katz. *Mastering Audio: The Art and the Science*. Butterworth-Heinemann, 2nd edition, 2014.

[17] Takuji Koike, Hiroshi Wada, and Toshimitsu Kobayashi. Modeling of the human middle ear using the finite-element method. *The Journal of the Acoustical Society of America*, 111(3):1306–1317, February 2002.

[18] Arthur L. Newcomb Jr. and Richard N. Young. Practical loudness: An active circuit design approach. *Journal of the Audio Engineering Society*, 24(1):32–35, 1976.

[19] Sofus Birkedal Nielsen. A loudness function for analog and digital sound systems based on equal loudness level contours. In *Audio Engineering Society Convention 140*. Audio Engineering Society, 2016.

[20] NIOSH-CDC. *Occupational Noise Exposure—Revised criteria for a recommended standard*. 1998.

[21] Harry F. Olson. The measurement of loudness. *Audio Magazine*, 56(2):18–22, 1972.

[22] Andrew J. Oxenham. Revisiting place and temporal theories of pitch. *Acoustical science and technology*, 34(6):388–396, 2013.

[23] Christopher J. Plack. *The sense of hearing*. Routledge, 2018.

[24] Torben Poulsen. Loudness of tone pulses in a free field. *The Journal of the Acoustical Society of America*, 69(6):1786–1790, 1981.

[25] Pradeep D. Prasad. A low complexity approach for loudness compensation. In *Audio Engineering Society Convention 129*, volume 1, pages 310–319. Audio Engineering Society, Nov. 2010.

[26] Ville Pulkki and Matti Karjalainen. *Communication acoustics: an introduction to speech, audio and psychoacoustics*. John Wiley & Sons, 2015.

[27] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, and S. M. Williams. *Neuroscience*. Sinauer Associates, 3^{rd} edition, 2008.

[28] Derek W. Robinson and R. S. Dadson. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7(5):166, 1956.

[29] Stuart Rosen and Peter Howell. *Signals and Systems for Speech and Hearing*. BRILL, 2^{nd} edition, April.

[30] Alan Seefeldt. Loudness domain signal processing. In *Audio Engineering Society Convention 123*. Audio Engineering Society, 2007.

[31] Julius O. Smith III. *Introduction to Digital Filters with Audio Applications*. W3K Publishing, 2007.

[32] Julius O. Smith III. *Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects*. W3K Publishing, 2010.

[33] Julius O. Smith III. Matched z-transformation. `https://ccrma.stanford.edu/~jos/pasp/Matched_Z_Transformation.html`, 2018.

[34] Stanley Smith Stevens. A scale for the measurement of the psychological magnitude: loudness. *American Psychological Association*, (43, No.5):405–416, 1936.

[35] Miguel Torres and Fernando Giráldez. The development of the vertebrate inner ear. *Mechanisms of development*, 71(1-2):5–21, 1998.

[36] Georgia Tsirimokou, Costas Psychalinos, and Ahmed Elwakil. *Design of CMOS analog integrated fractional-order circuits: applications in medicine and biology*. Springer, 2017.

[37] Vesa Välimäki and Joshua Reiss. All about audio equalization: Solutions and frontiers. *Applied Sciences*, 6(5):129, 2016.

[38] Stephen E. Widén, Sara Båsjö, Claes Möller, and Kim Kähäri. Headphone listening habits and hearing thresholds in swedish adolescents. *Noise & health*, 19(88):125, 2017.

[39] Jeremy M. Wolfe, Keith R. Kluender, and Dennis M. Levi. *Sensation & Perception*. Sinauer Associates Inc., 4th edition, 2014.

[40] William A. Yost. Psychoacoustics: a brief historical overview. *Acoustics Today*, 11(3):46–53, 2015.